Arts & Sciences Electronic Theses and Dissertations

Arts & Sciences

Summer 8-15-2013

Development of Polarizable Force Field Models for Transition Metal Ions

Jin Yu Xiang
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences

Computational and Molecular Biophysics

Dissertation Examination Committee:

Garland Marshall, Chair

Anders Carlsson

Cynthia Lo

Liviu Mirica

Jay Ponder

Scott Wildman

Development of Polarizable Force Field Models for Transition Metal Ions

by

Jin Yu Xiang

A dissertation presented to the

Graduate School of Arts and Sciences

of Washington University in

partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

August 2013

St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgements

The graduate study is a long and humbling learning process that would not be as enjoyable an experience if not for the people who have helped me along the way. First and foremost, I would like to thank my thesis advisor, Dr. Jay Ponder, for his guidance, support and encouragement without which this project would not be possible. His work on the AMOEBA force field is the foundation of my thesis and he has made important contributions to the field of computational chemistry and biology.

Another person I am especially indebted to is Dr. Garland Marshall, whom not only serves as my thesis committee chair, but also gave me my first experience in scientific research when I worked in his laboratory during my undergraduate study. Conversations with Dr. Anders Carlsson and Dr. Liviu Mirica have benefited me tremendously in broadening my understanding of the physics and chemistry of transition metal ions. I am also grateful for the insightful inputs from other members of my thesis committee members, Dr. Cynthia Lo and Dr. Scott Wildman.

In addition, I have learnt a great deal from the members of the Ponder lab, Dr. Mike Schneiders, Dr. Chuanjie Wu, Dr. Sergio Urahata and Dr. Yong Huang, and I am truly grateful for their friendship. I want to thank Dr. Pengyu Ren and Dr. Johnny Wu from University of Texas for sharing their findings on Zn(II) and in general their work on the AMOEBA force field. I also had the opportunity to rotate in Dr. Peter Chiver's and Dr. Nathan Baker's laboratories and I really value the experience I gained from them.

The Division of Biology and Biomedical Sciences, as well as the Department of Biochemistry and Molecular Biophysics, has provided me incredible exposure and

education to a wide range of scientific topics and I am very appreciative for that. I would like to express my gratitude to my supervisor at Genentech Inc., Dr. Vickie Tsui, for encouraging me to pursue my graduate degree. Last but not least, I have to thank all my family and friends for their support.

Dedicated to Mom and Dad,

Ya Bo Xiang and Hang Wa Wu

ABSTRACT OF THE DISSERTATION

Development of Polarizable Force Field Models for Transition Metal Ions

by

Jin Yu Xiang

Doctor of Philosophy in Computational and Molecular Biophysics

Washington University in St. Louis, 2013

Professor Garland Marshall, Chair

This dissertation focuses on the development of polarizable molecular mechanics (MM) force field models for the third-row transition metal (TM) ions. These TM ions perform important structural and chemical functions in a wide range of organic and biological environments because of the unique properties of the 3d orbitals. Being able study these systems *in silico* can provide a tremendous amount of information that is difficult to obtain through experiments. However, the standard treatment of ions in traditional MM models has shown to be insufficient for describing the d-shell electronic effects. In this work, empirical models for TM electronic effects are derived from the valence bond (VB) theory and the angular overlap model (AOM). The TM potential functions are incorporated into the AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) MM force field. A consistent polarizable electrostatics model is applied between metal and ligand sites at all interaction distance, enabling the study of ligand association / dissociation and other dynamic events. Specifically, theories are presented in the context of Ni(II), Cu(II) and Zn(II) ions. Parameters are obtained by fitting the TM models to gas-phase *ab initio* computations. Finally, results from molecular dynamics simulations of aqueous ions and select type 1 copper proteins

(plastocyanin and azurin) are analyzed. Evidence from this study suggests that explicit description of d-shell electronic effect can significantly improve the performance of MM models. This allows one to perform more reliable investigations on complex TM systems than can be achieved with traditional MM methods but without the computational expense of *ab initio* calculations.

Chapter 1. Introduction

Transition metal (TM) ions are involved in a tremendously diverse range of important biological systems.[1-7] Their functions can range from regulation of structure and binding, as in the zinc-finger motif in DNA recognition proteins[8], to active mediation of chemical reactions, as in the entire family of superoxide dismutase[9,10], cytochromes and electron transport proteins[11]. TM ions are also found in a number of pharmaceutical organometallic compounds, such as titanium, ruthenium and platinum-based antitumor drugs[3]. This incredible diversity in the applications of TM ions, both in natural biological systems and synthetic therapeutic vehicles, owes much to the unique properties of the valence d- and f-orbitals.[12] It is common that the coordination structures of TM ions are coupled to their chemical functions.[13] In the following sections, we demonstrate, via molecular orbital (MO) theory[14], the effects of d-orbitals on the ligation of the third-row TM ions. These ions are the most common TM species found in biological environments.[13] Existing quantitative computational techniques that can be used to study these TM systems are introduced, along with initial calculations that illustrate the various quantum effects.

1.1 Molecular Orbital Theory of the Third-Row Transition Metal Ions

Figure 1.1 is a graphical representation of the five unperturbed d-orbitals. It is clear from the figure that the d-orbitals have strong directionality and complex nodal features, which have significant implications for their interaction with TM ligands. In isolated atoms, the d-orbitals have equivalent ground state energies, but this redundancy is removed when the d-orbitals hybridize with overlapping ligand orbitals. The results of such hybridization depend on both the relative energy levels and orientation of the ligand orbitals.

Metal-ligand (M-L) interactions can be categorized into σ- ("head-on") and π- ("sideway") bonds. As a first approximation, the bonding and anti-bonding MOs of ligands can often be neglected since the energy differences relative to the d-orbitals are too large for significant hybridization to occur. For organic ligands, the non-bonding MOs that need to be considered when constructing the hybridized MOs are the valence s and p orbitals. Figure 1.2 shows schematically a few examples of how ligand MOs can participate in σ- and π- interactions. It is evident that the metal and ligand orbitals must have the correct symmetry for non-zero overlap to occur. σ-interactions are in general much stronger than π-interactions due to better geometric overlap between the interacting orbitals. Therefore, one can consider π-interactions as a perturbation to the principle ligand field constructed from the σ-interactions.

1.1.1 Molecular orbitals of octahedral ML$_6$ complexes

To construct the MO diagram under the effect of the principle ligand field for an octahedral TM complex with identical ligands, one needs to consider the shape, symmetry and energy ordering of the interacting orbitals. It is convenient to utilize group theory for this analysis. The character table of the O_h point group (Table 1.1) suggests that metal d-orbitals have E_g and T_{2g} symmetries. On the other hand, 2 e_g, 3 t_{1u} and 1 a_{1g} orbitals form the symmetry adapted linear combination orbitals (SALCO) of the ligands. The t_{2g} metal orbitals are therefore non-bonding, since they do not have any overlap with SALCO of the ligands. The e_g orbitals hybridizes to form two bonding and two anti-bonding MOs. The bonding MOs are concentrated on the ligands because they are usually more electronegative than the metal ion. Thus, the d-block valence MOs of the

3

octahedral TM complex consists of doubly degenerate e_g anti-bonding and triply degenerate t_{2g} non-bonding orbitals. Figure 1.3 illustrates the splitting of MOs from the ground state d-orbitals.

An important phenomenon for $Cu^{2+}L_6$ complexes that is directly related to the d-orbital splitting is the Jahn-Teller distortion.[15] Since Cu^{2+} has a d^9 electron configuration, three electrons have to be placed into two anti-bonding e_g MOs. Distortion along the molecular fourfold symmetry axis removes the degeneracy of these e_g MOs and thereby lowers the overall energy of the complex. This is commonly observed as an elongation of axial ligand bonding distance.

1.1.2   Molecular orbitals of square-planar and tetrahedral ML_4 complexes

The MO description of square-planar ML_4 complex can be readily deduced from the MOs derived for the corresponding octahedral ML_6 complex. If we take the ML_6 fourfold symmetry axis as the z-axis, removing the axial ligands has the effect of lowering the overlap between the metal d_{z^2} orbital and its interacting ligand orbitals. The resulting MO is still weakly anti-bonding due to non-zero overlap in the xy-plane (See Figure 1.1). The other MOs in the d-block remain unperturbed since they have zero coefficients along the z-axis. Therefore, we arrive at the MO diagram shown in Figure 1.4. Note that only two of the three non-bonding MOs are degenerate by symmetry.

We can follow a similar process to analyze the MOs of tetrahedral ML_4, starting from the results obtained for the square-planar geometry. If we place two ligands along both of the x-axis and y-axis, we can move from square-planar to tetrahedral coordination by pivoting the x-oriented pair above, and the y-oriented pair below, the xy-plane. In this

new orientation, the z^2 and xy d-orbitals are strictly non-bonding, since the ligands are now placed on the nodal planes of these orbitals. The yz and xz orbitals form anti-bonding MOs, as the ligands have moved out of their nodal planes. Finally, the MO derived from the $x^2 - y^2$ orbital has less anti-bonding character due to the fact that the ligand orbitals now overlap at an angle to the d-orbital. By consulting the character table of the T_d point group (Table 1.2), we can conclude that yz, xz and $x^2 - y^2$ form three degenerate anti-bonding MOs. The complete MO diagram can be found in Figure 1.5.

The MO description an ML_4 complex with square-planar ligand arrangement is very different from that for tetrahedral ligation. The square-planar geometry has three non-bonding, one weakly anti-bonding and one highly anti-bonding MO, whereas the tetrahedral coordination has two non-bonding MOs and three MOs with moderate anti-bonding energies. These differences in MO energy levels have dramatic impact on the geometric preferences of ML_4 complexes. For the d^{10} $Zn^{2+}L_4$ complex, tetrahedral coordination is preferred because the highest occupied molecular orbital (HOMO) is lower in energy than that in a square-planar arrangement. Low-spin $Ni^{2+}L_4$ complex (no unpaired electrons) strongly favors the square-planar geometry because only two electrons are anti-bonding, versus four in the case of tetrahedral coordination. However, the energy levels of the two geometries are reversed in the case of high-spin $Ni^{2+}L_4$ as a single electron has to be placed in the high-energy b_{1g} MO if the geometry is square-planar. The low-spin configuration is usually preferred since there are fewer anti-bonding electrons. Nevertheless, the tetrahedral geometry minimizes ligand-ligand repulsions and the exchange energy from two unpaired electrons favors the high-spin electronic state. The geometric preference for the intermediate d^9 $Cu^{2+}L^4$ complex is more nuanced. The

two coordination structures are very close in energy and the lowest energy arrangement depends on the identity of the ligands. In certain cases such as $[CuCl_4]^2$, it has been found that both structures can exist in equilibrium.[16] It is also clear from analyzing the MO diagrams that the ligation geometries are coupled to the redox potentials of TM ions, since changing d-block MO occupancies can significantly alter the energy levels of the coordination structures.

1.2 Quantum Mechanical Computational Methods

The qualitative analyses in the previous section have shown that the electronic configuration of TM complexes can have significant effect on their coordination geometries. A robust way to compute the energy for a TM system is by performing *ab initio* electronic structure calculations. Various relevant quantum mechanical (QM) computational methods are introduced in this section, largely following the notations of Jensen.[17]

1.2.1 Hartree-Fock method

To perform QM calculations on a static structure, one needs to solve the time-independent electronic Schrödinger's equation:

$$\mathbf{H}\Psi = E\Psi \tag{0.1}$$

An important approximation enabling us to solve the Schrödinger's equation is the Born-Oppenheimer approximation, in which the coupling of motion between nuclei and electrons is neglected. However, an exact solution is not available even in the context of this approximation, except for simple single electron species such as H_2^+. Instead, *ab*

initio numerical methods based on the variational principle are employed. Any trial wave function has an energy value greater than or equal to the exact solution and therefore one can variationaly improve the trial wave function by minimizing the energy. A Slater determinant (SD) is a convenient trial wave function that satisfies the antisymmetric property upon exchanging the coordinates of two electrons. For a general *N*-electron system, the SD takes the form:

$$\Phi_{\text{SD}} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \cdots & \phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) & \cdots & \phi_N(N) \end{vmatrix} \tag{0.2}$$

where ϕ denotes a one-electron orbital function. It can be shown that in the case of a single SD and non-relativistic Hamiltonian, the energy of the wave function can be written as:

$$E = \sum_i^{N_{\text{elec}}} \langle \phi_i | \mathbf{h}_i | \phi_i \rangle + \frac{1}{2} \sum_{ij}^{N_{\text{elec}}} \left(\langle \phi_j | \mathbf{J}_i | \phi_j \rangle + \langle \phi_j | \mathbf{K}_i | \phi_j \rangle \right) + V_{\text{nn}} \tag{0.3}$$

$\mathbf{h}_i$, $\mathbf{J}_i$ and $\mathbf{K}_i$ are operators that describe the one-electron motion due to the field of nuclei, the Coulombic repulsion between electron distributions ϕ_i^2 and ϕ_j^2, and the exchange energy that is a consequence of electrons being indistinguishable fermions respectively. The nuclear-nuclear repulsion V_{nn} does not depend on electron distribution and is therefore constant. The variation of the energy can be written in the form of a Fock operator $\mathbf{F}_i$:

$$\delta E = \sum_{i}^{N_{\text{elec}}} \left(\langle \delta\phi_i | \mathbf{F}_i | \phi_i \rangle + \langle \phi_i | \mathbf{F}_i | \delta\phi_i \rangle \right)$$

$$\mathbf{F}_i = \mathbf{h}_i + \sum_{j}^{N_{\text{elec}}} \left(\mathbf{J}_j - \mathbf{K}_j \right) \tag{0.4}$$

The orbital functions must be varied in an orthonormal manner. Using the Lagrange multiplier method, we have:

$$\delta L = \delta E - \sum_{ij}^{N_{\text{elec}}} \lambda_{ij} \left(\langle \delta\phi_i | \phi_j \rangle - \langle \phi_i | \delta\phi_j \rangle \right) = 0 \tag{0.5}$$

Solving this equality gives the final Hartree-Fock (HF) equations:

$$\mathbf{F}_i \phi_i = \sum_{j}^{N_{\text{elec}}} \lambda_{ij} \phi_j \tag{0.6}$$

These equations can be further simplified by applying a unitary transformation to the matrix of Lagrange multipliers such that $\lambda_{ij} = 0$ and $\lambda_{ii} = \epsilon_i$:

$$\mathbf{F}_i \phi_i' = \epsilon_i \phi_i' \tag{0.7}$$

where the ϕ_i' are known as canonical MOs. The HF orbital wave functions can only be determined when all the occupied orbitals are known. The overall procedure is therefore a self-consistent problem, and the solutions to the HF equations are the self-consistent field (SCF) orbitals. It should be noted that the use of a single SD approximation means that the electron correlation effect are ignored under the HF scheme. It is common to represent the orbitals by a set of basis functions:

$$\phi_i = \sum_{\alpha}^{M_{\text{basis}}} c_{\alpha i} \chi_\alpha \tag{0.8}$$

Substituting this into Equation 1.7:

$$\mathbf{F}_i \sum_\alpha^{M_{basis}} c_{\alpha i} \chi_\alpha = \epsilon_i \sum_\alpha^{M_{basis}} c_{\alpha i} \chi_\alpha \qquad (0.9)$$

Rewriting the result in matrix form, we obtain the well-known Roothaan-Hall equations for solving the HF equations, as implemented in many computer program packages

$$\mathbf{FC} = \mathbf{SC}\epsilon \qquad (0.10)$$

1.2.2   Electron correlation methods

For many polyatomic systems, ignoring electron correlation contributions can introduce significant error into energy calculations. The most straightforward approach to account for electron correlation is to simply remove the restriction to a single SD. This can be achieved by allowing electrons to occupy virtual orbitals and form excited SDs. The SDs can be singly (S), doubly (D), triply (T) *etc.* excited relative to the HF determinant. The trial wave function can then be taken as a linear expansion of ground and excited Slater determinants, with coefficients that allow energy to be at a minimum. This method based on the variational principle, is called configuration interaction (CI). A major drawback of this approach is that the number of excited Slater determinants grows factorially with the number of electron and basis functions. Therefore, full CI calculations are feasible only for small systems. Truncated CI, which takes into account excited SD up to a certain state, are usually used instead. Nevertheless, full CI is the best energetic result a particular basis set function can obtain in the limit of the Born-Oppenheimer approximation.

Alternatively, one can make use of the perturbation theory to improve upon the HF results. A popular method is the Møller–Plesset (MP) perturbation theory, which takes the zeroth-order Hamiltonian as the sum over the Fock operator:

$$E(MP0) = \sum_i^{N_{elec}} \epsilon_i \qquad (0.11)$$

where n in MPn denotes the order or the perturbation. Summing over the eigenvalues of the Fock operator counts the electron-electron repulsion twice. Therefore the appropriate first-order correction should be:

$$E(MP1) = -\frac{1}{2} \sum_{ij}^{N_{elec}} \left(\langle \phi_j | \mathbf{J}_i | \phi_j \rangle + \langle \phi_j | \mathbf{K}_i | \phi_j \rangle \right) \qquad (0.12)$$

Hence,

$$E(MP0) + E(MP1) = E(HF) \qquad (0.13)$$

Thus, the first correlation correction comes from the MP2 level:

$$E(MP2) = \sum_{i<j}^{occ} \sum_{a<b}^{vir} \frac{\left(\langle \phi_i \phi_j | \phi_a \phi_b \rangle - \langle \phi_i \phi_j | \phi_b \phi_a \rangle \right)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \qquad (0.14)$$

where electrons are doubly excited from occupied orbitals i and j to virtual orbitals a and b. This can be compared to CISD, where CI is accounted up to singly and doubly excited states. The main difference between CISD and MP2 is that CISD approaches the exact solution variationally and is therefore size intensive, *i.e.* the correction term is insensitive to the system size. MP2, on the other hand, is size extensive and the error term is relatively constant with respect to system size.

Another perturbation method related to the MP approach is the coupled cluster (CC) theory. In MP, a type of excitation (single, double… etc.) is corrected to a certain order (2, 3… etc.). In contrast, the CC method is to include all corrections to a type of excitation to an infinite order. Thus CCSD is equivalent to MP∞(SD) where all contributions from single and double excitations are accounted for. For the purpose of developing molecular mechanics model for transition metal ions presented in later chapters, the MP2 method is used for all single-point calculations due to its excellent accuracies and lower computational requirements when compared to the more expensive CISD and CCSD methods.[17]

1.2.3 *Density functional theory*

Another QM method widely used to compute the energy of a system is the density functional theory (DFT), which states that there is a one-to-one correspondence between the electron density (ρ) of a system and its energy. The system energy is divided into three parts: the kinetic energy (T), nuclear-electron attraction (E_{ne}) and electron-electron (E_{ee}) repulsion. Analogous to HF method, the electron-electron repulsion term should consist of a Coulombic and an exchange component, with correlation included implicitly. Contemporary DFT methods are based on Kohn-Sham (KS) theory, where the kinetic energy is split into a part that can be computed exactly and a small correction term. If a hypothetical system has non-interacting electrons, the exact kinetic energy functional for a single SD is:

$$T_{\mathrm{S}} = \sum_{i}^{N_{\mathrm{elec}}} \left\langle \phi_i \left| -\frac{1}{2}\nabla^2 \right| \phi_i \right\rangle \qquad (0.15)$$

Kinetic correlation from interacting electrons can be treated as a correction term to T_{S}, which is absorbed into the exchange-correlation energy term. The KS orbitals can be obtained through essentially the same SCF protocol used for the HF method. The remaining task is to obtain functionals that approximately describe the exchange-correlation correction.

Among the numerous published exchange-correlation functionals, those based on the generalized gradient approximation (GGA) have demonstrated the best performance for TM ions.[18-20] These functionals depends not only on the local electron density, but also on its first derivative. One of the most widely used DFT methods is B3LYP, which utilizes the three-parameter Becke model[21] that consists of the Becke 88 (B88) exchange[22] and the Lee, Yang and Parr (LYP) correlation functional [23]. B2PLYP is also a popular functional that has demonstrated superior performance to B3LYP for complex TM systems.[24] It is a mixture of B88, LYP with HF exchange and second order perturbation correction to the KS orbitals, analogous to MP2 method. An empirical long-range dispersion correction is often combined with B3LYP.[25]

1.2.4 Computational efficiency of quantum mechanics calculations

Although QM calculations treat the electronic structure problem rigorously, they are very computationally expensive. CPU time required for most SCF-based SD methods, including DFT, formally scales as the fourth power of the number of basis functions. Electron correlation methods such as MP2 scales at least as M_{basis}^{5}. Higher-level theory such as CISD and CCSD has a computational cost on the order of at least M_{basis}^{6}.[17]

Nevertheless, there are several strategies that can improve the computational efficiency of QM methods.

For methods in which the SCF procedure is the limiting step, linear-scaling techniques can be applied.[26-28] This is usually accomplished by partitioning large molecular systems into spatial segments and subsequently either approximating the long-range interactions, or using divide-and-conquer strategy to piece together the partitions. Another way to speed up QM calculations is by using semi-empirical methods.[29,30] The main approximation made in semi-empirical methods is that products of basis functions depending on electrons located on different atoms are neglected. Furthermore, only valence electrons are treated explicitly and minimum basis set functions are often used. Empirical parameters are then introduced to compensate for these approximation based on fitting to experimental data. These techniques allow single-point QM calculations to be performed on systems as large as 10,000 atoms.[31] However, ensemble sampling (Section 1.3) of such large systems using semi-empirical methods is still prohibitively expensive.

Alternatively, instead of describing the entire molecular system by QM, a mixed quantum mechanics / molecular mechanics (QM/MM) method can be used. QM/MM has become an especially common approach for studying metalloproteins.[18] A select region of the system (usually around the metal binding site) is treated via an electronic structure method, such as DFT or semi-empirical method. The rest of the molecule is handled by molecular mechanics (Section 1.3). For condensed phase simulation, implicit solvent models are typically used. This hybrid approach dramatically reduces the overall computational cost and enables meaningful ensemble sampling for thousands of atom

count, depending on the QM model used.[32,33] However, this technique is still much more computationally intensive than purely MM methods and the treatment of the QM-MM interface is known to be problematic.[34]

1.2.5 Valence bond theory

Although the MO theory is convenient for determining the molecular wave functions, it is often more qualitatively intuitive to describe a molecular system in terms of valency and bonding using ideas from the valence bond theory.[35,36] For a wave function Ψ, we can define a set of natural orbitals (NO) Θ that are given by:

$$\boldsymbol{\Gamma}\Theta_i = p_i\Theta_i \tag{0.16}$$

where $\boldsymbol{\Gamma}$ is the first order reduced density operator. The eigenvalues p_i are the occupancies of the eigenvectors Θ_i. If $\boldsymbol{\Gamma}$ is a one-electron operator represented as:

$$\boldsymbol{\Gamma}^A = N_{\text{elec}} \int \left| \Psi(1,2,...N_{\text{elec}}) \right|^2 d\mathbf{r}_2\, d\mathbf{r}_3...d\mathbf{r}_{N_{\text{elec}}} \tag{0.17}$$

we obtain a set of natural atomic orbitals (NAO) for a molecular system:

$$\boldsymbol{\Gamma}^A\Theta_i^A = p_i^A\Theta_i^A \tag{0.18}$$

At the dissociation limit, NAOs are natural orbitals of isolated atom, but they are also atomic orbitals with the highest occupancy in the molecular environment. In similar fashion, one can variationally search for natural bond orbitals (NBO) that are 1 or 2-center local orthonormal orbitals yielding the highest electron occupancies in Lewis-like bonding structures. The leading $N/2$ orbitals sorted by occupancy levels (for closed shell,

14

N for open shell) form the Lewis-type NBOs while the remaining orbitals span the entire basis and describe delocalization or resonance effects.

NBOs are constructed from natural hybrid orbitals (NHO) *h*, which are in turn a linear combination of NAOs:

$$h_A = \sum_i a_i \Theta_i^A \tag{0.19}$$

A 1-center lone pair orbital *n* is simply a single NHO:

$$n_A = h_A \tag{0.20}$$

A 2-center bonding NBO between atoms A and B is:

$$\Omega_{AB} = a_A h_A + a_B h_B \tag{0.21}$$

The corresponding anti-bonding NBO is:

$$\Omega_{AB}^* = a_A h_A - a_B h_B \tag{0.22}$$

Note that unlike anti-bonding orbitals in standard MO theory, anti-bonding NBOs represent unused valence shell capacity that leads to resonance stabilization.

For most TM complexes, it is necessary to describe the hypervalency of metal-ligand bonding via a 3-center-4-electron hyperbond:

$$A:B\text{-}C \leftrightarrow A\text{-}B:C \tag{0.23}$$

We can represent the NBOs of A:B-C by:

$$n_A = h_A$$
$$\Omega_{CB} = a_B h_B + a_C h_C \tag{0.24}$$
$$\Omega_{CB}^* = a_B h_B - a_C h_C$$

Similary for A-B:C we have:

$$n_C = h_C$$
$$\Omega_{AB} = a_A h_A + a_B h_B \tag{0.25}$$
$$\Omega_{AB}^* = a_A h_A - a_B h_B$$

The principle Lewis structure is therefore:

$$\Psi_{A:B-C} = (n_A)^2 (\Omega_{BC})^2$$
$$\Psi_{A-B:C} = (n_C)^2 (\Omega_{AB})^2 \tag{0.26}$$

The resonance stabilization of the Lewis structures is given by:

$$n_A \rightarrow \Omega_{BC}^* : \qquad \Delta E_{A:B-C}^{(2)} = -2 \frac{\langle n_A | \mathbf{F} | \Omega_{BC}^* \rangle}{\epsilon_{BC^*} - \epsilon_A}$$
$$\tag{0.27}$$
$$n_C \rightarrow \Omega_{AB}^* : \qquad \Delta E_{A-B:C}^{(2)} = -2 \frac{\langle n_C | \mathbf{F} | \Omega_{AB}^* \rangle}{\epsilon_{AB^*} - \epsilon_C}$$

where $\mathbf{F}$ is the Fock operator. This delocalization effect is the result of mixing A:B-C and A-B:C, each with its own appropriate resonance weight. This concept is the basis for our later development of the AMOEBA-VB model for TM ions, as presented in Chapter 2

1.3 Molecular Mechanics Computational Methods

Biological metalloproteins are large complex systems. For example, human copper-zinc superoxide dismutase (SOD3) is a homotetrameric protein that containing 222 amino acid residues with more than 5,000 heavy atoms.[37] Hemoglobin has two α and

two β chains each consisting of 141 and 146 residues, respectively, for a total of 574 amino acids.[38] Using QM methods to study these proteins, especially those with multiple metal binding sites, quickly becomes prohibitively expensive. Instead, MM is the method of choice for modeling these large systems.[39-43]

1.3.1 Force field model

In MM, atoms are the smallest fundamental units of the system. Electrons are no longer treated as separate particles with their own degrees of freedom. Instead, empirical energy terms that are functions of atomic coordinates are used to account for various electronic effects and parameters are fitted to sets of experimental or QM calculated data. Furthermore, nuclear motion is described by classical Newtonian mechanics. Given a set of empirical functions that describe the energy (U) of the system, the force on atom i can be computed by:

$$\mathbf{F}_i = -\nabla_{\mathbf{r}_i} U \tag{0.28}$$

Hence MM models are often called force field models. The standard force field potential energy term can be expressed as:

$$U_{\text{pot}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{torsion}} + U_{\text{cross-term}} + U_{\text{oop}} + U_{\text{vdW}} + U_{\text{elec}} \tag{0.29}$$

The first five terms are valence terms that describe the bonding interactions between atoms, while the last two terms are non-bonded potential functions.

U_{bond} usually takes the form of a Taylor expansion around an equilibrium bonding distance (r_{ij}^0) between two atoms i and j:

17

$$U_{ij}^{\text{bond}} = k_2(r_{ij} - r_{ij}^0)^2 + k_3(r_{ij} - r_{ij}^0)^3 + + k_4(r_{ij} - r_{ij}^0)^4 + \ldots \tag{0.30}$$

Since the zeroth and first order expansion terms are omitted, the energy and force at the equilibrium distance due to bonding is zero. It is clear this formulism does not have the correct behavior when $r_{ij} \to \infty$, as the energy becomes infinite instead of approaching zero. A Morse potential[44] is an alternative function that satisfies these limiting conditions:

$$U_{ij}^{\text{Morse}} = D(1 - e^{-a_{\text{Morse}}(r_{ij} - r_{ij}^0)^2}) - D \tag{0.31}$$

While the Morse potential exhibits correct dissociation behavior, MM is not the preferred model for studying bond breaking since it lacks electronic degrees of freedom. The Taylor expansion bond fomulation is more commonly used because it provides sufficient flexibility in fitting bond vibrational frequencies.

Similarly, the angle bend term (U_{angle}) also usually employs a Taylor expansion around an optimal angle (θ^0) formed between two bonds.

$$U_\theta^{\text{angle}} = c_2(\theta - \theta^0)^2 + c_3\text{bond},\theta(\theta - \theta^0)^3 + + c_4(\theta - \theta^0)^4 + \ldots \tag{0.32}$$

The torsional energy (U_{torsion}) is a four-body term describeing the rotational barrier along a bond. This is usually implemented as a Fourier series:

$$U_\omega^{\text{torsion}} = \sum_n V_n \left(1 + \cos(n\omega - \gamma)\right) \tag{0.33}$$

n controls the periodicity of the function, which typically ranges usually from 1 to 3. The phase-shift γ is often assigned 0 for odd n and π for even n to maintain achirality of the term

For sp^2-hybridized atoms, there is significant energetic penalty associated with moving the center atom away from the trigonal plane. This penalty is not fully captured by the torsional term and requires an additional out-of-plane bending energy term U_{oop}:

$$U_{\chi}^{\text{oop}} = a\chi^2$$
$$U_{d}^{\text{oop}} = ad^2$$

(0.34)

The out-of-plane term can be a function of either an out-of-plane bending angle (χ) or a distance (d) to the trigonal plane formed by the three attached atoms.

Cross-term energy ($U_{\text{cross-term}}$) function describes the coupling of the bonds, angles and torsions. The most commonly implemented term is stretch-bend coupling, but other cross-terms may also be used, depending on the force field model.

The van de Waals (vdW) term is used to describe the non-polar interactions between two non-bonded atoms. This energy function has the requirement to be repulsive at short distance, due to the explicit overlap of electron clouds, and asymptotically approach zero at long-range. There is also be a negative energy region surrounding the ideal vdW contact distance that accounts for dispersion attractions between atoms. One of the oldest functions for vdW interaction is the Lennard-Jones potential[45]:

$$E_{ij}^{\text{LJ}} = \varepsilon \left[\left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - \left(\frac{r_{ij}^0}{r_{ij}} \right)^{6} \right]$$

(0.35)

Alternative vdW functions also in use, such as the buffered 14-7[46] and the Buckingham type potentials[47].

Most standard force fields implement similar functions for the energy terms introduced thus far. The most important differentiation among the various models is the approach to describing electrostatic interactions. Force fields, such as CHARMM[48], OPLS[49], AMBER[50] and GROMOS[51] , use point partial charges to represent the electron distribution. These charges are often, but not necessarily, atom-centered. Fictitious charge sites can also be included at bonded centers or other off-atoms sites to improve the model accuracy. The interaction between two point partial charges (q_i and q_j) is described by a simple Coulomb potential:

$$E_{ij}^{\text{elec}} = \frac{q_i q_j}{\varepsilon r_{ij}} \tag{0.36}$$

A more sophisticated model to describe the electrostatic potential around a molecule is to include contributions from higher order electric moments, such as dipole, quadrupole etc. These higher order moments are usually obtained by performing multipole analysis on MO wave functions derived from high-level QM calculations.[52] Example of force fields that employ a multipole electrostatic model are AMOEBA[53] and SIBFA[54].

Another crucial aspect of electrostatic interaction is polarization. In traditional force fields, the coupling between the local environment and the charge/multipole model is neglected. Multi-body electrostatic contributions are not accounted for, which can be significant for polar molecules.[55] A simple method to include polarization is by allowing partial charges to adjust to changes in molecular environment based on electronegativity equalization.[56] However, this first approximation approach is unable to capture charge polarization on planar molecules when the electric field is perpendicular to the molecular plane. Drude oscillator methods, also known as "shell models", have also been used to

account for polarization effects in force fields.[57] Polarization is introduced by attaching massless Drude particles to atoms via harmonic springs, which are allowed to move in response to the electric field. Alternatively, one can compute the induced dipole moment ($\mu_{i,\alpha}^{\text{ind}}$) explicitly in response to external field ($E_{i,\alpha}$):

$$\mu_{i,\alpha}^{\text{ind}} = \alpha_i E_{i,\alpha} \tag{0.37}$$

where α_i is the atomic polarizability.[58] This is the model implemented in the AMOEBA force field.

1.3.2   Molecular mechanics simulation techniques

Experimental measurement of an observable (O) of a molecular system is a time-average value of that property. The instantaneous value of O depends on the momenta **p** and positions **r** of all particles. Thus the ensemble average of O is given by:

$$\langle O \rangle = \int\int d\mathbf{p}\, d\mathbf{r}\, O(\mathbf{p},\mathbf{r})\rho(\mathbf{p},\mathbf{r}) \tag{0.38}$$

where ρ is the probability of a system with momenta **p** and positions **r**. Therefore, one must sample the conformational space of the molecular system with the correct probability distribution in order to correlate computed properties with experimental measurements. This is especially important for complex systems with multiple local minima that are close in energy. For the canonical ensemble (constant number of particle, volume and temperature), ρ is given by:

$$\rho(\mathbf{p},\mathbf{r}) = \frac{1}{Q_{\text{NVT}}} e^{-\frac{E(\mathbf{p},\mathbf{r})}{k_{\text{B}}T}}$$

$$Q_{\text{NVT}} = e^{-\frac{A(\mathbf{p},\mathbf{r})}{k_{\text{B}}T}}$$

(0.39)

where k_{B} is the Boltzmann constant, E is the energy of the system, Q_{NVT} is the partition function and A is the Helmholtz free energy.

One method to sample the conformational space is through Metropolis Monte Carlo (MMC) simulation. A trial conformational or configurational move is generated along a Markov chain and the difference in energy between the new and old structures is weighted by the Boltzmann factor. The trial move is only accepted if a uniform random number on [0,1] is lower than the Boltzmann factor. Alternatively, one can perform molecular dynamics (MD) calculations to simulate the time-evolution of the molecular system. As described previously, the forces on an atom can be readily computed from the negative gradient of the force field energy. Assuming these forces, and the corresponding accelerations, are constant for a very small time period (approximatly 1fs), a new set of new positions and velocities can be computed from the original conformation to produce an MD step. This stepping procedure is continued to produce a full MD trajectory of the system across time. Both MMC and MD are widely used sampling methods. The advantage of MD is that time-dependent properties can be easily computed from MD trajectories whereas there is no time relationship between two trial moves in MMC. Both methods automatically generate conformational probability densities in accordance to Equation (0.39). Hence the ensemble average is simply the arithmetic mean of observable the O in the generated MMC Markov chain or the MD trajectory generated.

1.3.3   Force field for transition metal ions

A number of different MM models have been reported that describe TM-ligand interactions with varying degree of success. The simplest approach is fitting traditional force field terms such as bonds, angles and torsions to known properties obtained from experiments or QM calculations. However, the force field parameters obtained through this process generally have limited transferability and different parameters may become necessary for the same type of ligand depending on ligation geometry. For example, the d^9 $Cu^{2+}L_6$ complex often has elongated axial ligands due to the Jahn-Teller distortion. This phenomenon is not well described if the same bond parameters are used for all the ligands.[43] More importantly, standard angular potentials based on a Taylor expansion of a reference ligand-metal-ligand (L-M-L) values or a Fourier series are inappropriate for describing TM complexes. ML_5 complexes such as $Fe(CO)_5$ adopts the trigonal bipyramidal geometry, where the angles between ligands can be 90, 120 and 180 degrees.[59] Another example is $[CuCl_4]^{2-}$, for which both square-planar and tetrahedral structures exist in equilibrium. (see Section 1.1.2)

A more radical solution is to construct a "reactive" model that allows atoms to respond chemically to their environment by dynamically assigning bond orders and charges based on molecular geometries.[60,61] Alternatively, there are "semi-classical" models that employ potential functions for TM ions derived from the valence bond (VB) theory[35,62-65] or the angular overlap model (AOM)[66] to supplement traditional force field energy terms. Models such as VALBOND[67-70] are based on a simplified version of the VB theory, in which TM ions are treated as hypervalent resonance centers and L-M-L interactions are described by geometric overlap between sd^n hybridized bonding metal-

ligand orbitals. On the other hand, models proposed by Deeth, *et al.*[71-73], Piquemal, *et al* [74], and Carlsson, *et al.*[75-78] are developed from the AOM and the ligand field (LF) effects are handled through explicit diagonalization of a perturbed d-orbital matrix due to the presence of ligands. These methods have demonstrated satisfactory agreements with experiments and with *ab initio* calculations when used to study a range of TM systems with different coordination geometries and ligation states.

The semi-classical force fields introduced thus far have focused on modeling the effects of local metal-ligand binding on the geometry of TM complexes. However, electrostatic interactions are also a major component of TM complex energetics. In most TM models, the electrostatic potential is not applied between metal and its ligands, which makes these inappropriate for study of ligand exchanges and other dynamic events. In addition, TM ions behave similarly to main group cations at distances beyond direct ligation, and polarization becomes an important contributing factor. Most semi-classical models developed for TM ions use a fixed charge model for electrostatic interactions, which is inadequate for treating systems with highly polar sites.[79] The main motivation for our current work is an attempt to address these shortcomings. It has been shown that the AMOEBA force field has excellent performance for main group mono- and di-cations[80,81] and therefore provides an appropriate basis for modeling TM di-cations.

1.4 Preliminary Investigations

To demonstrate the importance of d-orbital electronic effect on the coordination chemistry of TM complexes, we computed the energies of square-planar and tetrahedral $[M(NH_3)_4]^{2+}$ complexes at varying metal-ligand distances using MP2/6-311G(d,p) [82] QM

methodology. The metal species M are dications of chromium, manganese, iron, cobalt, nickel, copper and zinc. The metal spin states were chosen to give the lowest QM energies. The differences in potential energy between the two ligand geometries are plotted in Figure 1.6.

It is immediately clear that Cr^{2+}, Ni^{2+} and Cu^{2+} adopt a square-planar geometry when ligated to NH_3 while other third-row TM di-cations prefers the tetrahedral geometry. Since all these ions have formal +2 charges and have similar atomic radii,[83] standard force field treatment of ions will not be able to correctly describe these geometric preferences. Another consequence of the d-orbital splitting arises in the subtlety of spin states. For example, the d^6 Fe^{2+} complex at first glance should be low-spin, in which all three non-bonding d-block orbitals of the square-planar geometry are doubly occupied. (Section 1.1.2) However, QM calculations reveal that the high-spin tetrahedral structure is favored, indicating that the exchange energy has compensated for the ligand field effect.

1.5 Figures
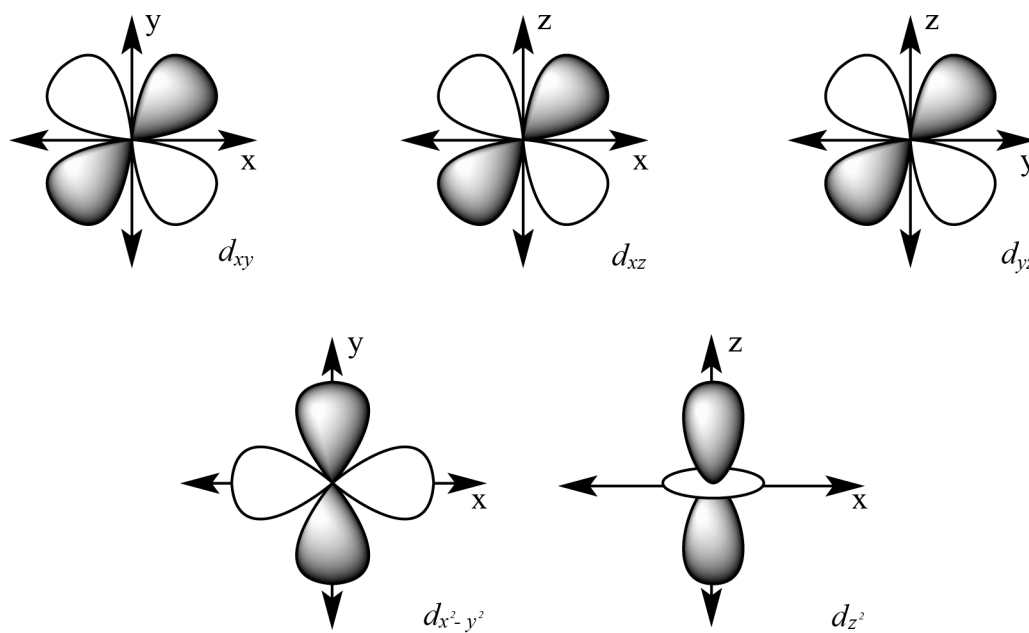
Figure 1.1 Graphical representations of ground state d-orbitals
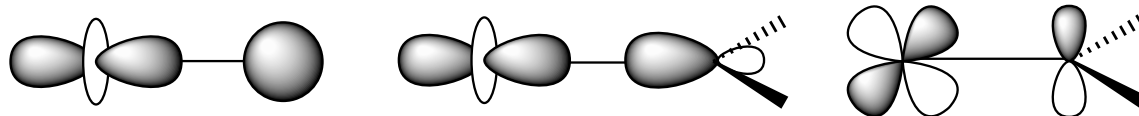
Figure 1.2 Illustration of sample σ- (left and middle) and π- (right) bonds
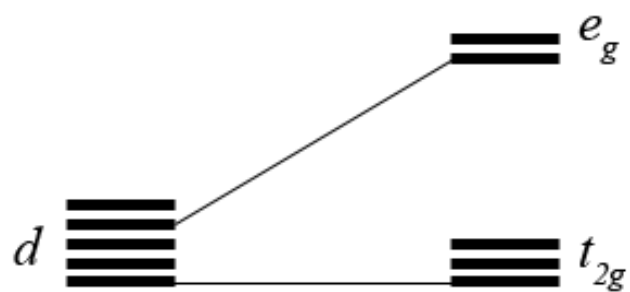
Figure 1.3 MO diagram demonstrating the d-orbital splitting for octahedral ML_6 complex
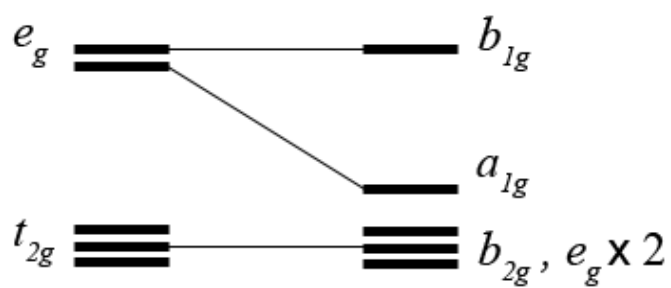
Figure 1.4 MO diagram showing the relationship between ML_4 orbitals (right) and ML_6 orbitals (left)
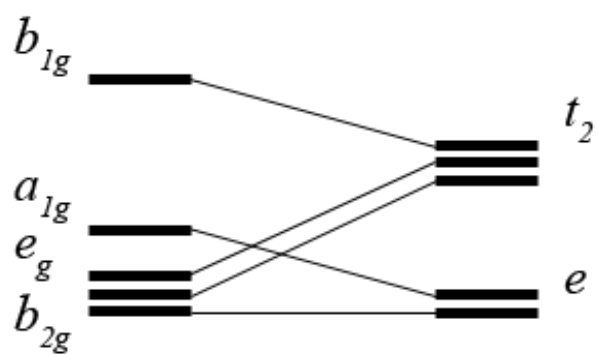
Figure 1.5 MO diagram showing the relationship between square-planar ML$_4$ orbitals (left) and tetrahedral ML$_4$ orbitals (right)
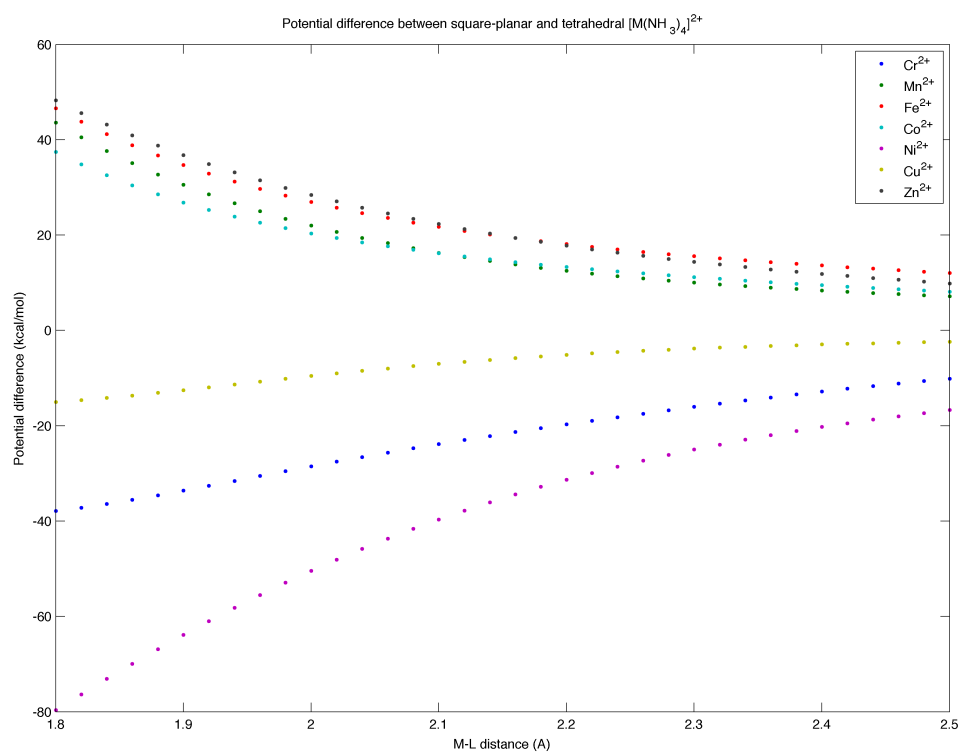
Figure 1.6 Energy difference between square-planar and tetrahedral tetra-aqua $[M(NH_3)_4]^{2+}$ complexes computed using MP2/6-311G(d,p) at varying metal-ligand separations; energy calculated by subtracting the potentials of tetrahedral from that of square-planar structures

1.6 Tables

Table 1.1 Character table for the O_h point group

O_h	E	$8C_3$	$6C_2'$	$6C_4$	$3C_2$	i	$8S_6$	$6\sigma_d$	$6S_4$	$3\sigma_h$	
A_{1g}	1	1	1	1	1	1	1	1	1	1	$x^2 + y^2 + z^2$
A_{2g}	1	1	-1	-1	1	1	1	-1	-1	-1	
E_g	2	-1	0	0	2	2	-1	0	0	2	$(z^2, x^2 - y^2)$
T_{1g}	3	0	-1	1	-1	3	0	-1	1	-1	
T_{2g}	3	0	1	-1	-1	3	0	1	-1	-1	(xy, xz, yz)
A_{1u}	1	1	1	1	1	-1	-1	-1	-1	-1	
A_{2u}	1	1	-1	-1	1	-1	-1	1	1	-1	
E_u	2	-1	0	0	2	-2	1	0	0	-2	
T_{1u}	3	0	-1	1	-1	-3	0	1	-1	1	(x, y, z)
T_{2u}	3	0	1	-1	-1	-3	0	-1	1	1	

Table 1.2 Character table for the T_d point group

T_d	E	$8C_3$	$3C_2$	$6S_4$	$6\sigma_d$		
A_1	1	1	1	1	1		$x^2+y^2+z^2$
A_2	1	1	1	-1	-1		
E	2	-1	2	0	0		$(2z^2-x^2-y^2, x^2-y^2)$
T_1	3	0	-1	1	-1		
T_2	3	0	-1	-1	1	(x, y, z)	(xy, xz, yz)

33

Chapter 2. A Valence Bond Theory in the AMOEBA

Polarizable Force Field

In this chapter, a MM model is developed for aqueous Ni^{2+}, Cu^{2+} and Zn^{2+} ions based on VB theory in conjunction with the AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) polarizable force field.[58] The development of VALBOND by Landis, *et al.* suggests that VB theory may be incorporated into MM through relatively simple algebraic functions that are computationally efficient. In this initial investigation, we limit our scope to Ni^{2+}, Cu^{2+} and Zn^{2+} in order to reduce the number of spin states and the complexity of model development. Unless otherwise stated, we constrain discussions in this chapter on Ni^{2+} ion to its low-spin species. Parameters are determined against energies calculated with QM methods for metal-water complexes in the gas phase and validated against experimental data for the aqueous ions. Additionally, previous work shows that the AMOEBA force field provides a satisfactory description for the aqueous Zn^{2+} ion.[84] We have pursued further investigation to see if modeling the covalency explicitly between water ligands and Zn^{2+} can improve the accuracy of the existing AMOEBA model. In the following sections, we present the AMOEBA-VB framework for Ni^{2+}, Cu^{2+} and Zn^{2+} ions and document the procedures for obtaining force field parameters. Results from energy computations for gas phase ion-water complexes and molecular dynamics simulations for aqueous ion solutions are reported and compared against QM and previously published data.

2.1 Methodology

2.1.1 AMOEBA-VB framework

The general interatomic AMOEBA potential energy can be expressed as:

$$U_{\text{AMOEBA}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{b-a}} + U_{\text{oop}} + U_{\text{torsion}} + U_{\text{vdW}} + U_{\text{ele}}^{\text{perm}} + U_{\text{ele}}^{\text{ind}} \qquad (1.1)$$

where the first five terms represent bond stretch, angle bend, bond-angle cross-term, out-of-plane bend and torsion potentials used to describe local valence contributions. The last three terms handle nonbonded interactions, including the van de Waals (vdW), permanent electrostatic and induced electrostatic potentials.[58,85,86] Additional potential energy terms for TM centers based on VB theory are added to the overall energy:

$$U_{\text{total}} = U_{\text{AMOEBA}} + U_{\text{VB}} \tag{1.2}$$

In the context of aqueous TM ions, only the nonbonded interactions from the standard AMOEBA model are applied between the metal center and water molecules.

2.1.2   Nonbonded intermolecular potentials

The basic AMOEBA potential terms use energy expressions from previous published reports.[58,85,86] A buffered 14-7 potential [46] is used to model vdW interactions, and takes the following form:

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

where $\rho_{ij} = R_{ij} / R_{ij}^0$, and R_{ij} represents the separation between atoms i and j. The values of n, m, δ and γ are set to 14, 7, 0.07 and 0.12 respectively, while ϵ_{ij} and R_{ij}^0 correspond to the potential energy well-depth and minimum energy distance. For heterogeneous atom pairs, mixing rules are applied to determine ϵ_{ij} and R_{ij}^0:

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

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

(1.4)

The electrostatic potential is described as having a permanent and an induced component. The permanent electrostatic component is represented by atom-centered monopole, dipole and quadruple moments. The parameters are determined via Stone's distributed multipole analysis[87] followed by refinement against QM-derived electrostatic potential maps. Polarization is accounted for via self-consistent induced dipoles computed from:

$$\mu_{i,\alpha}^{ind} = \alpha_i E_{i,\alpha}$$

(1.5)

where α_i is the atomic polarizability and $E_{i,\alpha}$ is the total electric field generated by permanent multipoles and induced dipoles. A Thole damping factor is applied at short interaction distances, corresponding to use of a smeared charge representation that takes the form:

$$\rho = \frac{3a}{4\pi} e^{-au^3}$$

(1.6)

where a is a dimensionless factor controlling the strength of damping and $u = R_{ij}/(\alpha_i\alpha_j)^{1/6}$ is the effective separation between polarizable sites i and j. The Thole mechanism serves to avoid the well-known polarization catastrophe at small separations,[88] and yields reasonable anisotropic molecular polarizabilities starting from isotropic atomic polarizability values.[86]

2.1.3 Water model

The AMOEBA water model has been previously reported,[85] and tested in a variety of different environments.[89] The standard intermolecular and intramolecular energy terms for water are retained in the AMOEBA-VB model. The water force field parameters for the nonbonded and valence potentials are reported in Table 2.1 and Table 2.2, respectively.

2.1.4 Transition metal ion model

In addition to the usual AMOEBA vdW and electrostatic potentials, VB terms are applied between each TM ion and ligand water oxygen atom, as water molecules interact with TM ions predominately through lone pair p-orbital electrons on the oxygen atoms. As a first approximation, a TM-water complex is modeled by its principle field, with water interacting with TM ions through σ bonding only.[14] The VB component is expressed as:

$$U_{VB} = \sum_k^{resonance} W_k U_{resonance,k} \tag{1.7}$$

where the total VB potential is the summation of individual energy contributions from the resonance structures corresponding to the TM complex. W_k is an empirical function that mimics the weighting for resonance structure k in natural resonance theory.[35]

For Ni^{2+}, Cu^{2+} and Zn^{2+} water complexes, the principle resonance structure corresponds to the Lewis structure, as shown in Figure 2.1a, where the TM interacts with water molecules via ionic interactions. The intermolecular energy of the principle

resonance structure corresponds to the regular AMOEBA non-bonded potentials. The d-electron effect can then be explained by considering minor non-Lewis resonance species where 3-center-4-electron (3c4e) bonds are formed between the TM center and ligand atoms.[35] This represents the donation of electron density from oxygen to the metal, and delocalized ionic-covalent bonding stabilizes the hypervalent TM center. Using this description, the molecular orbitals of Ni^{2+} and Cu^{2+}-water complexes can be decomposed into the contributions from Lewis and non-Lewis resonance structures. Note that using a single 3c4e bond per resonance is only valid for low-spin Ni^{2+}, which has an empty instead of two partially filled d-orbitals. Its 3c4e bonds have predominantly d character since the Ni^{2+} and Cu^{2+} 3d valence orbitals can accept electron density more readily than the 4s orbital. On the other hand, the 3d orbitals of Zn^{2+} are fully filled and the resonance hybrids are mainly due to overlap with the Zn^{2+} 4s orbital. Hence both Ni^{2+} and Cu^{2+} have greater resonance stabilization energy than Zn^{2+}. The overall hypervalent resonance scheme for the TM ions is shown in Figure 2.1b. The angle formed by a 3c4e bond will be referred to henceforth as the "resonance angle".

The intermolecular energy between a TM ion and ligand water molecules for an individual resonance construct k can be expressed as:

$$U_{\text{resonance},k} = U_{\text{VB-bond},k} + U_{\text{VB-angle},k} \qquad (1.8)$$

where $U_{\text{VB-bond},k}$ and $U_{\text{VB-angle},k}$ are the two bonding terms and one angular term used to describe a single 3c4e bond. Since the number of resonance structures is equal to the number of angles formed by the TM-water complex, the overall energy contribution from the VB component becomes:

$$U_{VB} = \sum_{k}^{angles} W_k (U_{VB\text{-}bond,k} + U_{VB\text{-}angle,k}) \tag{1.9}$$

VB angular potential is based on Pauling's principle of angular overlap for a pair of $sp^m d^n$ hybrid orbitals.[62,63,65] The overlap integral associated with the presence of two identical non-orthogonal $sp^m d^n$ bonding orbitals is:

$$\Delta = \sigma^2 + \pi^2 \cos\theta + \frac{\delta^2}{2} (3\cos^2\theta - 1)$$
$$\sigma^2 = \frac{1}{1+m+n}, \ \pi^2 = \frac{m}{1+m+n}, \ \delta^2 = \frac{n}{1+m+n} \tag{1.10}$$

where θ is the angle between the orbitals. The terms σ^2, π^2 and δ^2 represent the s, p and d contributions to the bond, respectively. Following Landis,[68] we construct the angular potential for a 3c4e bond as:

$$U_{VB\text{-}angle,k} = K_{VB\text{-}angle,k} \left(1 - \Delta(\theta_k + \pi)^2\right) \prod_{i}^{2} F_{VB\text{-}angle,k,i}$$
$$F_{VB\text{-}angle,k,i} = e^{-\alpha_{k,i} r_{k,i}^2} \tag{1.11}$$

where $K_{VB\text{-}angle,k}$ is a constant scaling factor for angle k. The bond order term in Landis' formulation is folded into $K_{VB\text{-}angle,k}$ in our implementation. We introduce an additional scaling factor, $F_{VB\text{-}angle,k,i}$, as a function of the metal-ligand distance $r_{k,i}$ in bond i, and an empirical parameter $\alpha_{k,i}$. This factor is necessary to describe the overlap drop-off with increasing metal-ligand distance. The overall energy term has a linear geometrical preference that is suitable for describing 3c4e bonding involving Ni^{2+} and Cu^{2+}. The angular potential is not applicable to Zn^{2+} since the interacting 4s orbital is spherically symmetric. Previous data has shown AMOEBA satisfactorily describes aqueous Zn^{2+}

ions without the addition of a potential term to account for d-electron effects.[84] However, we retain the bonding component for Zn^{2+} to investigate its impact on the AMOEBA model. Note the result from overlapping hybrid orbitals is destabilizing and therefore the VB angular term is always positive. A Gaussian-like function is adapted for the VB bonding potential:

$$U_{\text{VB-bond},k} = -\sum_{i}^{2} K_{\text{VB-bond},k} F_{\text{VB-bond},k,i}$$
$$F_{\text{VB-bond},k,i} = e^{-\beta_{k,i} r_{k,i}^2}$$

(1.12)

where the index i sums over the two ligands in a single 3c4e hypervalent bond. $K_{\text{VB-bond},k,i}$ is the scaling parameter for bond i of resonance angle k. In contrast to the angular term, the VB bonding contribution is purely stabilizing. Additionally, we propose an empirical resonance weighting function for resonance structure k that is based on metal-ligand distances:

$$W_k = \prod_{i}^{2} F_{\text{resonance},k,i} \sum_{l}^{\text{angles}} \left(c_l + \prod_{j}^{2} F_{\text{resonance},l,j} \right)$$
$$F_{\text{resonance},k,i} = e^{-\gamma_{k,i} r_{k,i}^2}, \ F_{\text{resonance},l,j} = e^{-\gamma_{l,j} r_{l,j}^2}$$

(1.13)

where c_l is a parameter for resonance angle l. The index l runs through all resonance angles including k. The subscripts i and j denote the two metal-ligand pairs in resonance angles k and l, respectively. According to this formulation, the weighting for resonance construct k depends on the positions of all water molecules in the TM complex. Note that although the resonance weight function depends on the number of ligands, it is general for all coordination number and its value transitions smoothly between them.

Finally, it has been shown that Cu^{2+} complexes in octahedral geometries exhibit Jahn-Teller type distortions.[15,90,91] Since the simplified AMOEBA-VB model presented does not compensate for this effect explicitly, we explored the effect of adding a harmonic first order component[92] where the Jahn-Teller stabilization energy arises from the Q_θ distortion mode. The exact formulation used is:

$$E_{JT}^{xy} = -(r - r_0)\Delta / r_0$$
$$E_{JT}^{z} = -2(r - r_0)\Delta / r_0$$

(1.14)

where r is the metal-ligand distance, r_0 is the average bonding distance of the TM complex, and Δ is an empirical value to scale the strength of the Jahn-Teller effect. E_{JT}^{xy} and E_{JT}^{z} are applied to the in-plane and axial ligand molecules respectively.

2.1.5 *Parameterization and validation*

The parameters for the AMOEBA-VB framework are based on fitting MM energy values to those obtained by *ab initio* methods for structural variants derived from common ligation geometries of TM complexes, including square-planar, tetrahedral and octahedral. These structures are generated in such way that they represent easily accessible states during computational simulations. All electronic structure calculations were performed with the Gaussian 09 package.[93] QM geometry optimizations were carried out with B3LYP[21,23] DFT calculations using the 6-311G(d,p)[82] basis set. Single-point energy were computed via MP2/aug-cc-pVTZ [94] on main group elements and MP2/cc-pVTZ [95] for the TM ions. An SCF convergence criterion of 10^{-9} a.u. was imposed, and a Fermi-broadening SCF method[96] was used for Cu^{2+} complexes to improve convergence stability. The AMOEBA-VB potentials and Cartesian derivatives

were implemented in the TINKER[58] molecular modeling package used for all MM computations.

The $[M(H_2O)_4]^{2+}$ and $[M(H_2O)_6]^{2+}$ gas phase complexes were optimized using QM methods with angular constraints to yield idealized tetrahedral, square-planar and octahedral ligation geometries. Intramolecular optimization within water molecules was allowed. These structures serve as a starting point for generating further variations in geometry designed to assess different aspects of the MM model. Complex energies computed by AMOEBA were manually fit to QM data from corresponding procedures using a common set of parameters for a metal ion interacting with a single ligand molecule. The standard AMOEBA parameters were optimized, and then fixed in value, prior to fitting the VB terms. Results with the Jahn-Teller distortion term were also computed when applicable. $[Ni(H_2O)_6]^{2+}$ is not included in this initial model development as it is a high-spin species that would necessitate a different resonance formulation. As a result, aqueous simulations for the Ni^{2+} ion were not performed because it has been suggested the Ni^{2+} first solvation shell consist of six water molecules.[97]

Bond stretching. Starting from optimized structures with idealized bonding geometries for square-planar $[Ni(H_2O)_4]^{2+}$ and $[Cu(H_2O)_4]^{2+}$, tetrahedral $[Zn(H_2O)_4]^{2+}$ and octahedral $[M(H_2O)_6]^{2+}$, single point energy calculations were performed with both QM and MM methods and plotted as a function of varying metal-oxygen distance (see Figure 2.2a). Water molecules were held rigid during this procedure. The protocol was designed to test the accuracy of the MM model in describing bonding potentials for ideal ligation geometries.

43

Hypervalent effect. Without accounting for the resonance effect due to the hypervalent center, gas phase metal-water complexes adopt geometries that minimize ligand-ligand repulsion.[43] Hence main group tetra-aqua complexes favor a tetrahedral geometry over the corresponding square-planar configuration. The presence of strong 3c4e resonance hybrids for Ni^{2+} and Cu^{2+} is predicted to stabilize the square-planar geometry according to VB theory. On the other hand, the lack of an angular contribution from 3c4e bonding for Zn^{2+} leads it to prefer a tetrahedral water complex. Therefore, the energetic difference between tetrahedral and square-planar structures provide a direct indication of the magnitude of the hypervalent effect. Single point energies were computed by QM and MM methods for $[M(H_2O)_4]^{2+}$ in both square-planar and tetrahedral coordination, and at varying metal-oxygen distances. All water molecules were kept equidistance from the TM center for each data point (see Figure 2.2b). Energy differences between the two geometries, after removing the water-water interaction energy in the absence of a metal ion, are calculated and plotted with respect to the metal-oxygen separation.

Random perturbation. We use a series of perturbed metal-ligand structures to gain insight into whether the MM model can reproduce the *ab initio* energy surface near the optimized structures. Small random perturbations were introduced to optimized ideal geometries by changing the metal-ligand distances and rotating the ligand around the metal-ligand vector and two orthogonal axes (Figure 2.2c). The maximum perturbation from the optimized structure was 0.2Å for metal-ligand distance and 10 degrees for each rotation. Structures containing ligand-ligand contact distances less than 2.5Å were discarded, and a total of 100 random complex geometries were generated. The energy of

each complex was computed by QM, and compared against values obtained from MM models. Structures with QM energies more than 15 kcal/mol higher than that of the idealized geometry were discarded since these high-energy structures are not readily accessible during routine MD simulations.

Molecular dynamics. Molecular dynamics simulations were performed for both aqueous Cu^{+2} and Zn^{+2} ions using the parameters derived above. A total of 8ns of canonical ensemble MD trajectory at 298K was collected for a single TM ion and 214 water molecules in a 18.6216 Å cubic box. Periodic boundary conditions were applied and particle-mesh Ewald summation was utilized to include long-range electrostatic interactions.[98,99] The convergence criterion for self-consistent dipole polarization was set to a 0.01 Debye RMS change in atomic induced dipole moments. The correlation function, solvation shell properties and coordination number of each TM ion was computed from the trajectories and compared to published data.

2.2 Results and Discussions

2.2.1 Energy components

The values for parameters obtained from the fitting procedures are shown in Table 2.3. The TM ions are assigned only a +2 permanent charge; it does not make sense for TM ions to possess higher-order multipoles in the absence of an external electric field. The polarizability and Thole damping factor are similar to those of main group dications in previously published studies.[54,81] The vdW radii follow the general trend across third row transition metals in that $Zn^{2+} \geq Cu^{2+} \geq Ni^{2+}$.[83]

45

The Ni^{2+} and Cu^{2+} VB parameters are obtained with the 3c4e bond hybridization set to 10% s and 90% d (corresponding to $\sigma^2 = 0.1$, $\pi^2 = 0$ and $\delta^2 = 0.9$ in Equation (1.10)). We obtained this empirical ratio by recognizing that oxygen lone-pair electrons predominately interact with d orbitals of the Ni^{2+} and Cu^{2+} ions, which have d^8 and d^9 configuration respectively in their ground states. A small amount of s hybridization is modeled to take into account the effect of d-s mixing. Figure 2.3 shows the overall shape of the VB angular potential, which is similar to the corresponding function derived by Carlsson, *et al.*[75,76] from AOM considerations. The main features of the potential function are the two local minima at ligand-metal-ligand angles of 180° and 90°, allowing tetra-aqua Ni^{2+} and Cu^{2+} complexes to adopt the preferred square-planar geometry.

The QM optimized metal-ligand distances for tetra- and hexa-aqua TM complexes are reported in Table 2.4. For tetra-aqua complexes, all four water molecules remain equidistance from the TM center, after bond relaxation under symmetry angular constraints. However, the axial and basal water molecules for hexa-aqua Cu^{2+} complexes adopt very different ligation distances as a result of Jahn-Teller distortion.[91] The axial water molecules in $[Cu(H_2O)_6]^{2+}$ are significantly elongated, and this presents a challenge for MM models lacking separate parameters for axial and basal water molecules as shown in the results below. The AMOEBA-VB energy breakdown for these optimized geometries is presented in Table 2.5. Note the VB bonding and angular components are reported in conjunction with resonance weighting as this reflects the final energy contributions from both 3c4e interactions and resonance as indicated in Equation (1.9). In terms of relative strength of the various energy components, the permanent electrostatic

46

interaction makes the largest individual contribution, followed in order by the polarization, vdW and VB potential energies.

2.2.2 *Bonding potential*

Bonding potential energies computed by QM and MM methods are shown in Figure 2.4. A single bond potential is constructed for tetra-aqua complexes since the water molecules are equidistant from the metal. However, axial and basal water molecules for octahedral complexes are plotted separately due to their differences in bonding distances and energies. In the cases of $[Ni(H_2O)_4]^{2+}$ and $[Cu(H_2O)_4]^{2+}$, both the AMOEBA and the AMOEBA-VB models arrive at minimum energy distances consistent with QM values, but the inclusion of the VB components produces a stronger binding interaction that better reflects QM results. For $[Cu(H_2O)_6]^{2+}$, both MM models produce the correct bonding geometry for basal water molecules, with AMOEBA-VB again producing a more accurate interaction energy. Neither model was able to reproduce the full extent of the elongation of axial ligand to metal distances, resulting in 2.07Å and 2.09Å for AMOEBA and AMOEBA-VB respectively versus 2.33Å for QM. The interactions between axial water molecules and the Cu^{2+} ion are also too strong (-23.07 kcal/mol for AMOEBA-VB, -24.23 kcal/mol for AMOEBA and -18.07 kcal/mol for QM), in general agreement with the distance discrepancies. Adding an explicit Jahn-Teller distortion term does not dramatically improve the ligand binding geometry (axial Cu-O distance at 2.12Å) but it does produce a more accurate binding energy (-17.94 kcal/mol). Results from the MM model with and without the VB term do not exhibit a significant difference for tetra- and hexa-aqua Zn^{2+} complexes. For Zn^{2+}, both MM methods produce bonding potentials in agreement with QM calculations.

2.2.3 Hypervalent effect

The energy difference between gas phase square-planar and tetrahedral tetra-aqua complexes are plotted as a function of metal-ligand distance in Figure 2.5. Note that water-water interactions are subtracted to isolate the energetics between TM and water molecules. It is apparent from the figure that in the absence of a VB component, AMOEBA produces the wrong geometrical preference for $[Ni(H_2O)_4]^{2+}$ and $[Cu(H_2O)_4]^{2+}$. The AMOEBA-VB framework is able to capture the correct trend of the hypervalent effect, even though the computed energy difference is still relatively small compared to QM data. As our final proposed model, we have settled on a set of parameters producing the most balanced performance across all aspects of the parameterization. Figure 2.5 also suggests the VB angular potential is not required to obtain the optimal tetrahedral geometry for $[Zn(H_2O)_4]^{2+}$ complex.

2.2.4 Energy surface

To help assess the accuracy of the MM energy surface, we compare in Figure 2.6 the energies computed using *ab initio* methods with those from MM for perturbed structures around idealized geometries. All energy values presented are relative to the energy of idealized coordination structures. Results obtained with the AMOEBA-VB framework show there is a dramatic 60% and 18-19% reduction in RMS deviation from QM values when compared with AMOEBA-only data for Ni^{2+} and Cu^{2+} complexes respectively. Addition of the Jahn-Teller distortion term does not materially change the results. On the other hand, the addition of the VB term to Zn^{2+} does not have a meaningful impact on correlation between QM and MM results. For these species, both

AMOEBA and AMOEBA-VB are able to generate accurate relative potential energies in comparison with QM data.

2.2.5   Ions in aqueous solution

A series of canonical ensemble molecular dynamic simulations were performed for aqueous solutions containing a single Cu^{2+} or Zn^{2+} ion. Calculation for Cu^{2+} used the AMOEBA-VB model, but without application of the Jahn-Teller distortion term. Omission of the Jahn-Teller was necessary during MD because the simple first harmonic potential function does not provide a smooth energy transition when axial and basal ligands rearrange during the course of a simulation. The metal-oxygen correlation function and radial distribution function for water surrounding the TM ion is presented in Figure 2.7. The first solvation shell for both TM ions is found to contain six water molecules and the ligation geometries, along with data from previous studies, are reported in Table 2.6. Six-membered ligation states have been reported in the literature for Zn^{2+} [97,100,101] and this agrees with our observation. However, there is a lack of general consensus regarding the optimal ligation geometry of aqueous Cu^{2+}, and a variety of first solvation shell occupancies have been reported.[102,103] A solvation number of 5-6 has been suggested for Cu^{2+} from numerous experimental and computational studies. [97,104-106] The 5-coordinate structure is generally attributed to a distortion from octahedral geometry due to the Jahn-Teller effect. We did not observe the "dual-peak" 6-coordinate Cu-O radial distribution obtained from simulation with the ReaxFF model.[61]

2.3 Conclusions

The AMOEBA-VB framework presents a foundation upon which a generalized transition metal force field can be built. The appeal of a MM model based on VB is that it is physically intuitive and avoids differential treatment of ligands of the same type based solely on coordination geometry. The results presented show addition of VB components to AMOEBA improves energetic accuracy when compared to QM data, while producing reasonable simulation results in aqueous solution. It is also clear that AMOEBA can satisfactorily describe the characteristics for aqueous Zn^{2+} without explicit modeling of the interaction between oxygen lone-pair electrons and TM orbitals.

2.4 Figures

a

Principle Resonance

b

Hypervalent Resonance
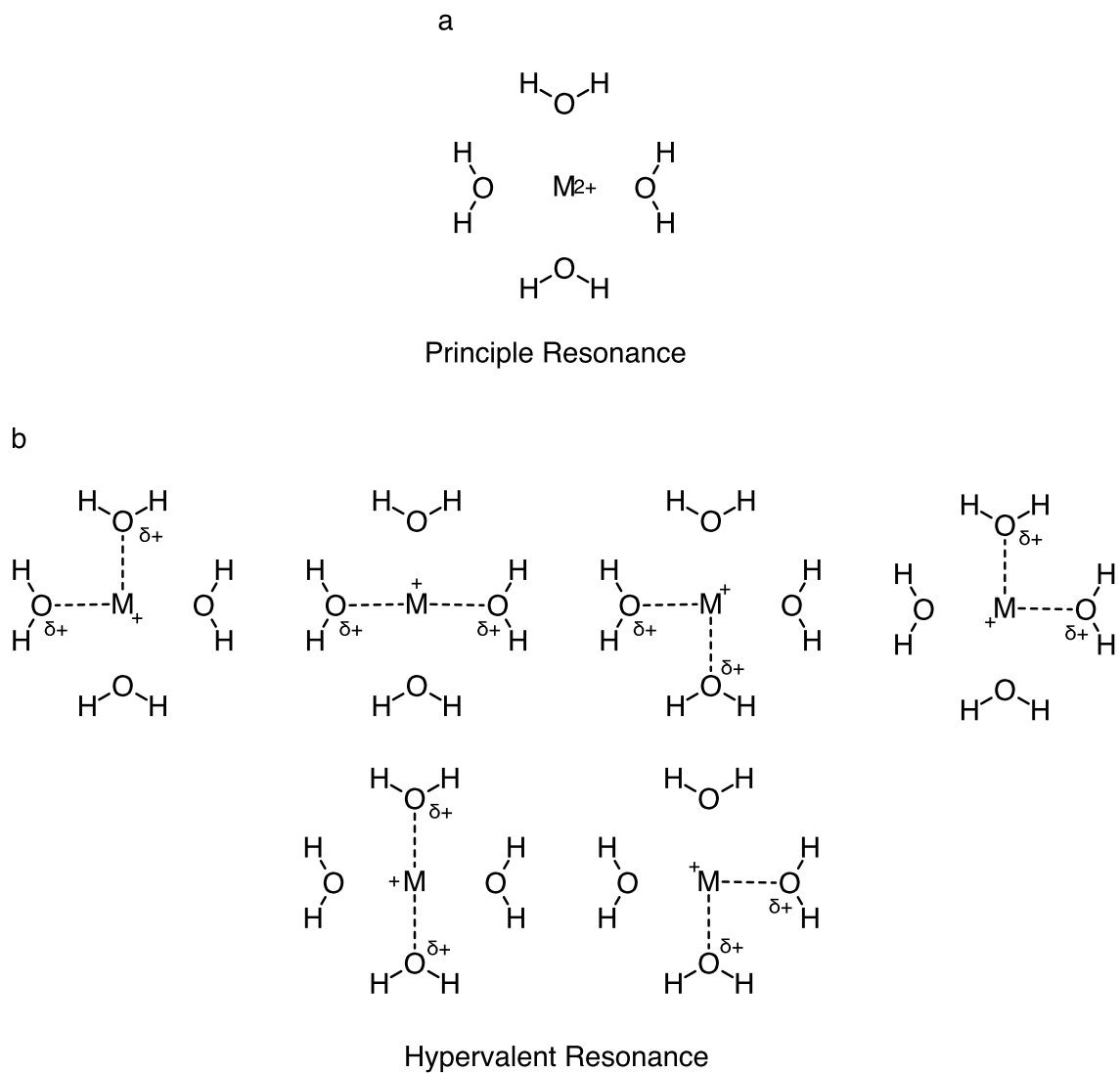
Figure 2.1 Resonance scheme for $[M(H_2O)_4]^{2+}$ complex where M = Cu or Zn. a) Principle resonance that corresponds to the Lewis structure of the complex. b) Non-Lewis minor hypervalent resonance structures with a single 3c4e bond per resonance; the number of such resonance structures is equal to C_2^n where n is the number of ligands.

a. bond stretch

b. ligand field stretch

c. random perturbation

water

TM

stretch

rotation
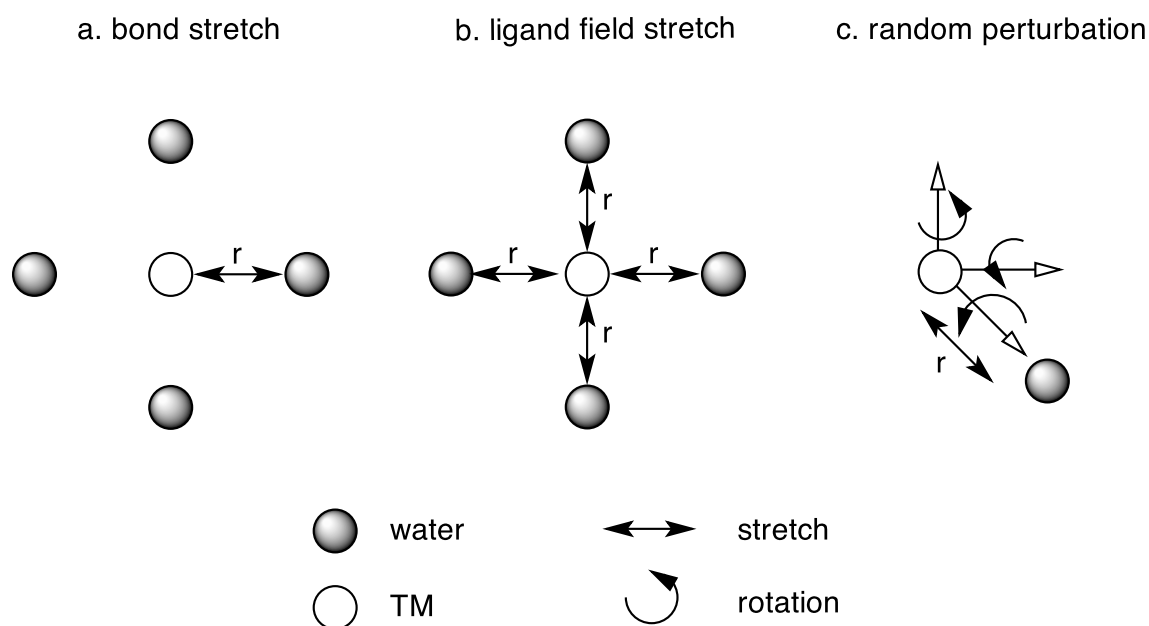
Figure 2.2 Methods for generating TM complex structural variations from idealized geometries used in AMOEBA-VB and QM gas phase calculations. a) a single TM-ligand distance is varied while other ligands are fixed at their QM-optimized coordinates. b) all TM-ligand distances are changed simultaneously from the optimized geometry and each ligand remains equidistance to the metal center during the process. c) perturbations are introduced to TM-water complexes by randomly changing the metal-ligand distances and rotating around the local metal-ligand vector and two axes orthogonal to the vector.

$$1 - \Delta(\theta + \pi)^2$$
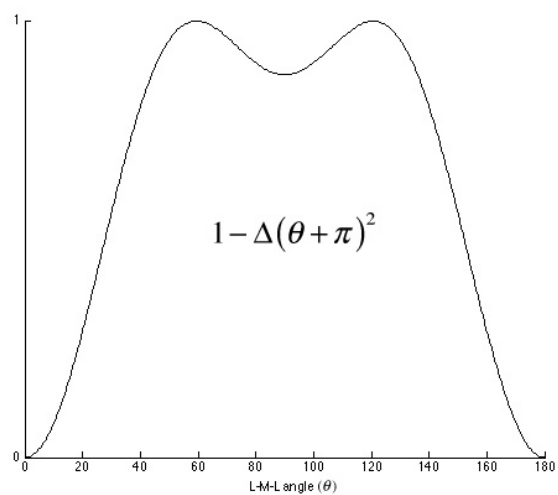
Figure 2.3 Schematic plot of VB angular potential for each 3c4e bond based on 10% s and 90% d hybridization.
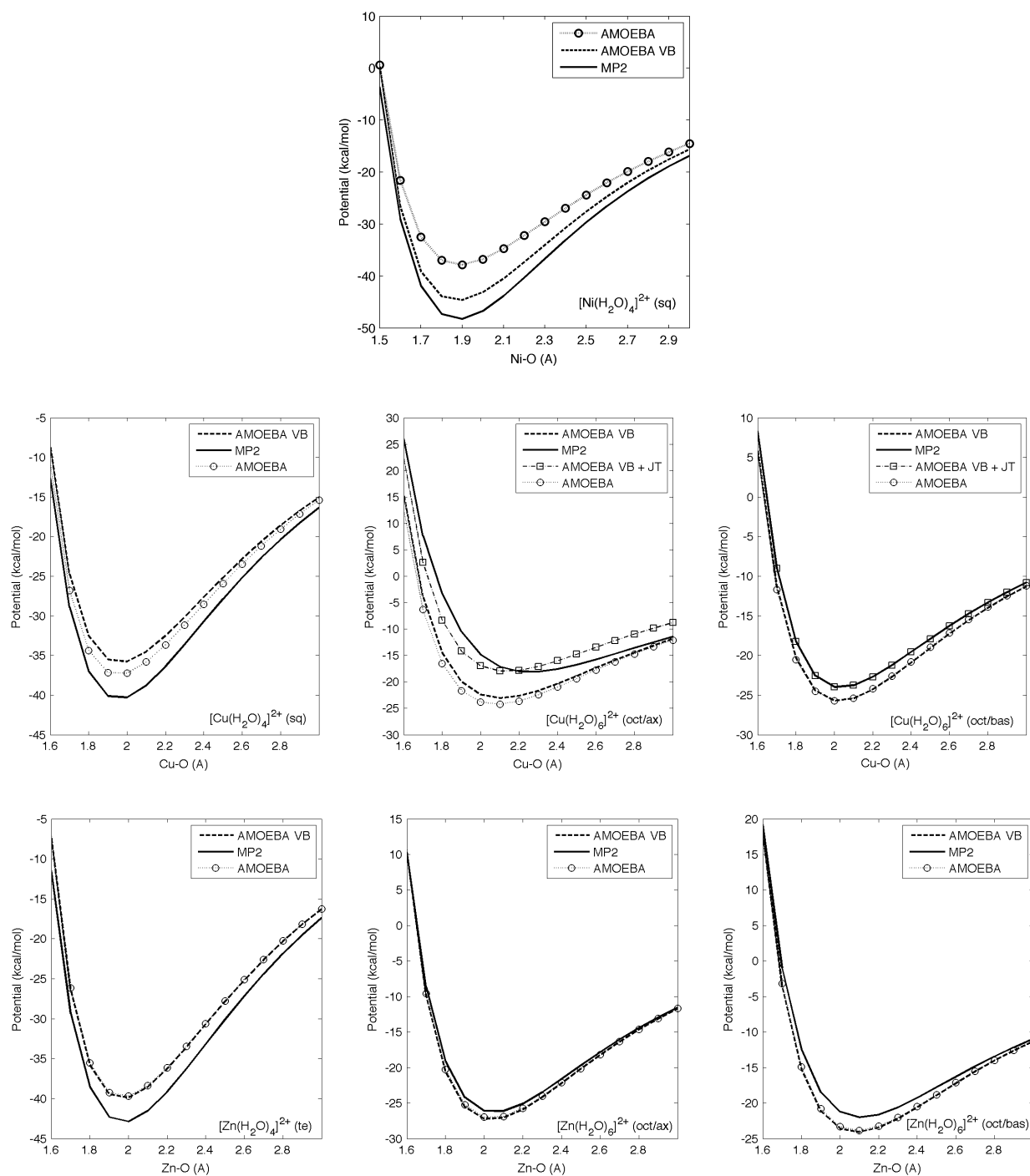
Figure 2.4 Comparison of bond potentials between QM and MM methods; the zero potential is set as the energy of a complex at 5Å metal-oxygen separation in order to approximate dissociation; see supporting information Table 1 and 2 for numerical values. Abbreviations: sq = square-planar, te = tetrahedral, oct = octahedral, ax = axial, bas = basal, JT = Jahn-Teller distortion term.
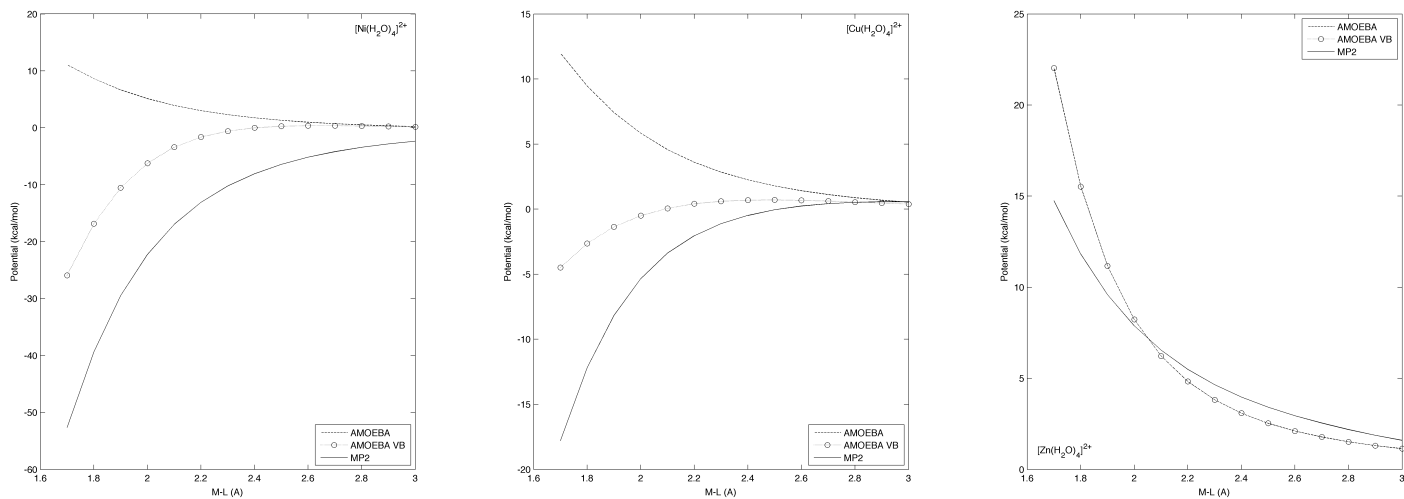
Figure 2.5 Energy difference between square-planar and tetrahedral tetra-aqua TM complexes; energy calculated by: $\left(U_{sq} - U_{sq/empty} \right) - \left(U_{te} - U_{te/empty} \right)$; data points from AMOEBA and AMOEBA-VB methods for $[Zn(H_2O)_4]^{2+}$ overlap each other since the differences in results are very small.
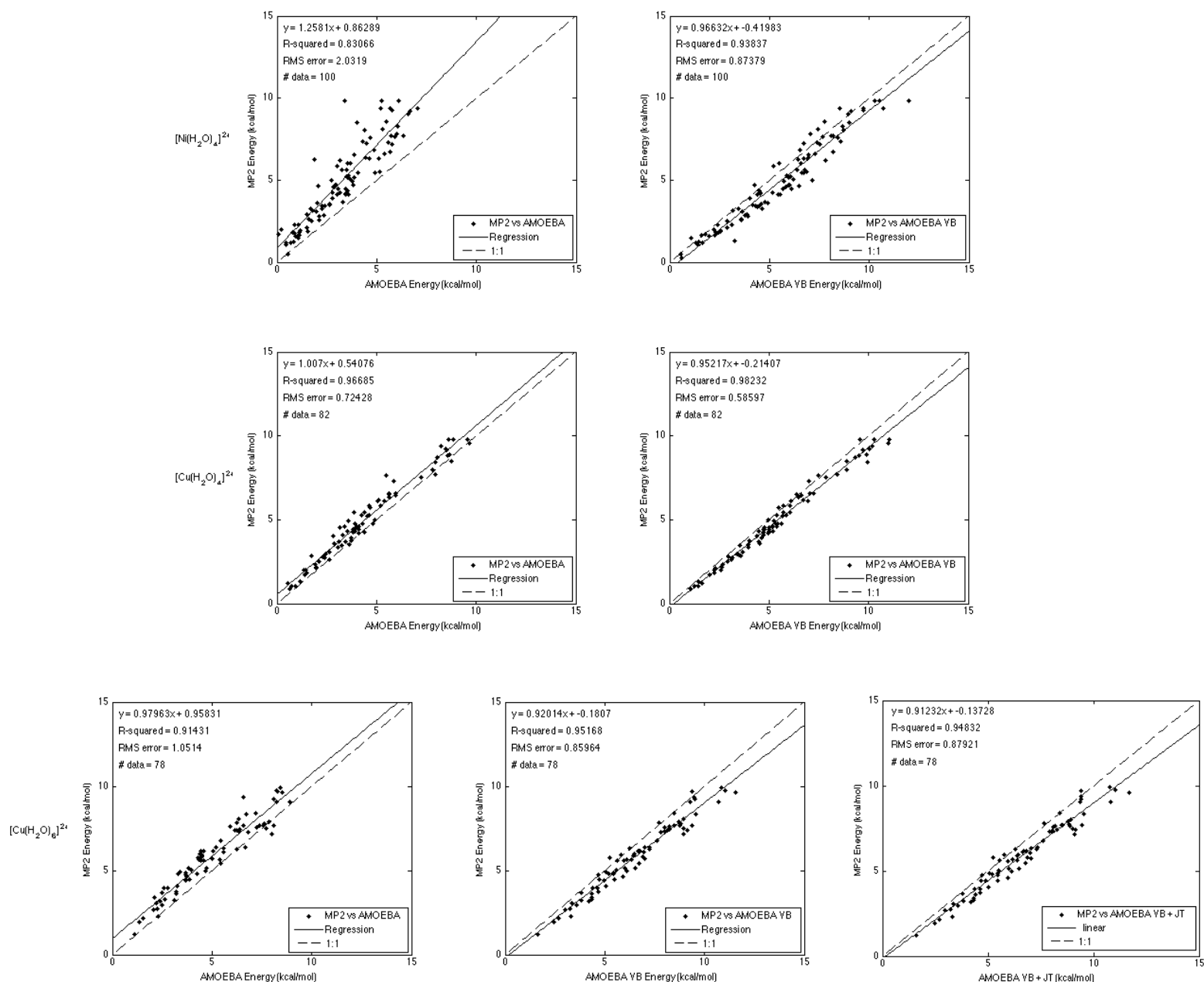
Figure 2.6a Comparisons of QM and MM energies for perturbed aqua Ni^{2+} and Cu^{2+} structures; for $[M(H_2O)_4]^{2+}$, results without VB components are on the left and that with VB terms are on the right; for $[Cu(H_2O)_6]^{2+}$, results without VB term, with VB term, with VB and Jahn-Teller distortion terms are plotted in the left, middle and right panel respectively.
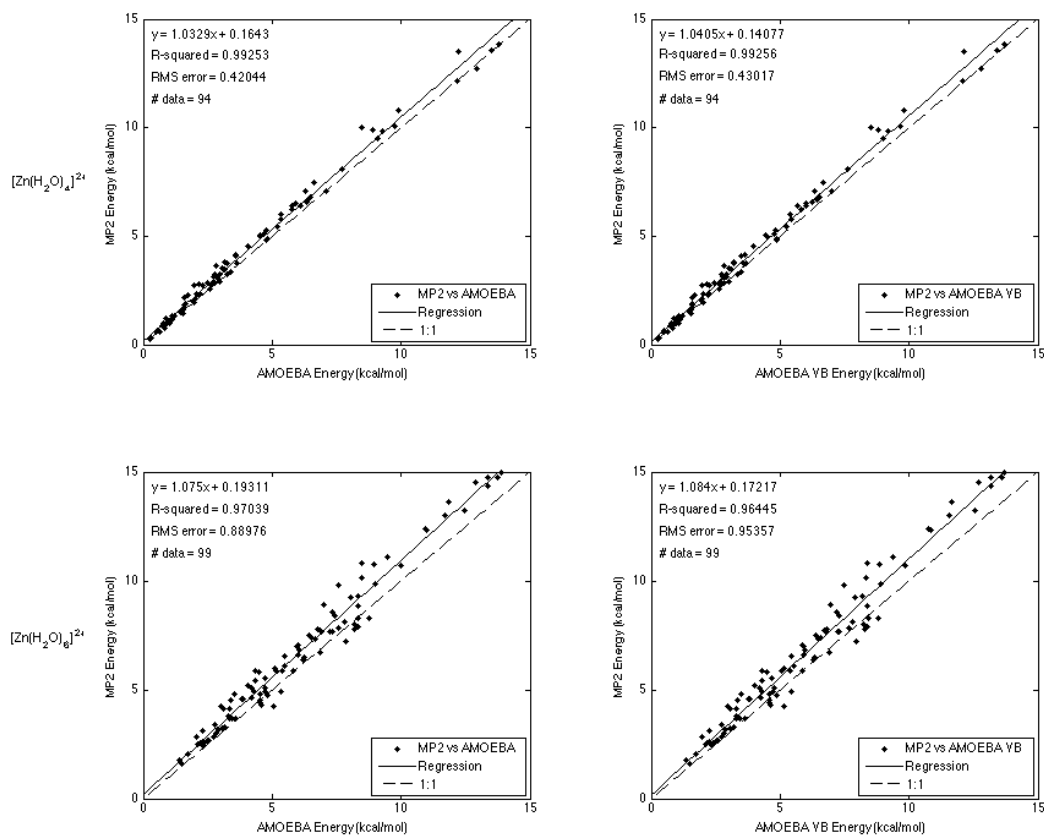
Figure 2.6b Comparisons of QM and MM energies for perturbed aqua Zn^{2+} structures; results without VB components are on the left and those with VB terms are on the right.
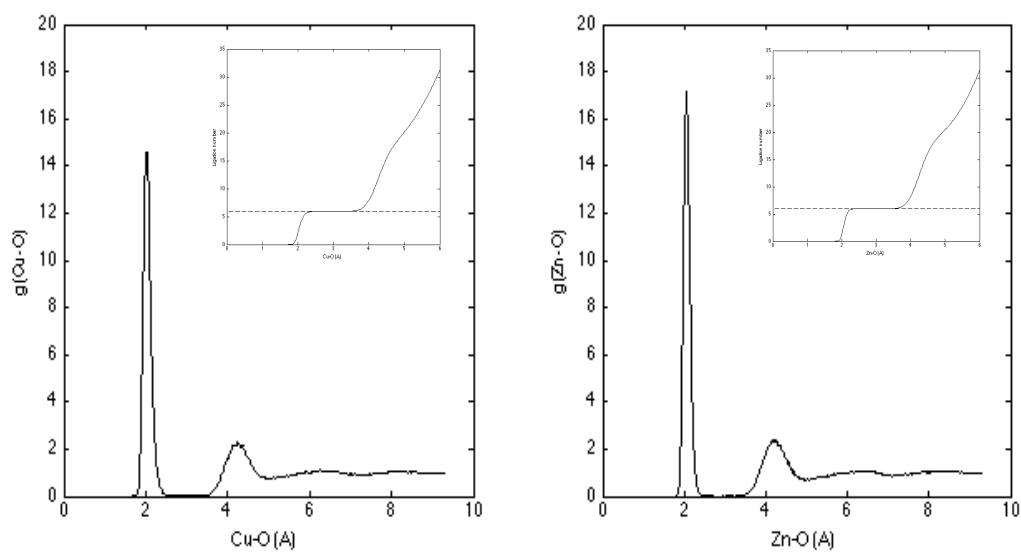
Figure 2.7 Metal-oxygen correlation function and radial distribution of water molecules surrounding a TM center (insert). The dashed line corresponds to a first solvation shell with six water molecules.

2.5 Tables

Table 2.1 Intermolecular (vdW and electrostatic) potential parameters for AMOEBA water; $*M=[q,\mu_1,\mu_2,\mu_3,Q_{11},Q_{12},...,Q_{33}]^{\mathrm{T}}$.

	ϵ (kcal/mol)	R_0 (Å)	$M*$ (a.u.)				α (Å^3)	a
O	0.1100	3.405	$\begin{bmatrix} -0.51966 & \cdots & & \\ 0.00000 & 0.00000 & -0.14279 & \cdots \\ 0.37928 & 0.00000 & 0.00000 & \cdots \\ 0.00000 & -0.41809 & 0.00000 & \cdots \\ 0.00000 & 0.00000 & 0.03881 & \end{bmatrix}$				0.837	0.39
H	0.0135	2.655	$\begin{bmatrix} 0.25983 & \cdots & & \\ -0.03859 & 0.00000 & -0.05818 & \cdots \\ -0.03673 & 0.00000 & -0.00203 & \cdots \\ 0.00000 & -0.10739 & 0.00000 & \cdots \\ -0.00203 & 0.00000 & 0.14412 & \end{bmatrix}$				0.496	0.39

Table 2.2 Intramolecular potential parameters for AMOEBA water.

Potential	Force Constant	Ideal Length/Angle
Bond Stretching	K_b = 529.6 kcal/mol/Å^2	b_0 = 0.9572 Å
Angle Bending	K_θ = 34.05 kcal/mol/radian^2	θ_0 = 108.5˚
Urey-Bradley	K_l = 38.25 kcal/mol/Å^2	l_0 = 1.5537 Å (H...H)

Table 2.3 Force field parameters for TM ions; the second row for each TM ion represents values fitted without VB component; *TM ions are assigned only permanent monopole equal to their formal charge; [#]only applied to octahedral complexes.

	vdW		Electrostatics			VB-bond		VB-angle		Resonance		Jahn-Teller
	ϵ	R^0	$M*$	a	α	$K_{\text{VB-bond}}$	$a_{\text{VB-bond}}$	$K_{\text{VB-angle}}$	$a_{\text{VB-angle}}$	a_{res}	b_{res}	$\Delta^{\#}$
Ni^{2+}	0.34	2.88	2	0.16	0.05	1.820	0.27	1.625	0.24	0.39	140.0	--
	0.15	2.80	2	0.18	0.19	--	--	--	--	--	--	--
Cu^{2+}	0.24	2.88	2	0.16	0.12	5.49	0.20	9.01	0.22	0.22	2.00	0.40
	0.24	2.88	2	0.16	0.12	--	--	--	--	--	--	--
Zn^{2+}	0.34	2.90	2	0.16	0.12	0.20	0.30	--	--	0.30	10.00	--
	0.34	2.90	2	0.16	0.12	--	--	--	--	--	--	--

Table 2.4 Metal-ligand distances in B3LYP/6-311G(1d,1p) optimized geometries for tetra- and hexa-aqua Ni^{2+}, Cu^{2+} and Zn^{2+} gas phase complexes.

$[Ni(H_2O)_4]^{2+}$ (sq)	4×1.91Å
$[Cu(H_2O)_4]^{2+}$ (sq)	4×1.93Å
$[Zn(H_2O)_4]^{2+}$ (te)	4×1.98Å
$[Cu(H_2O)_6]^{2+}$ (oct)	4×2.03Å $+ 2 \times 2.33$Å
$[Zn(H_2O)_6]^{2+}$ (oct)	4×2.10Å $+ 2 \times 2.16$Å

Table 2.5 AMOEBA-VB energy breakdown for QM optimized complex geometries; Energy values are in kcal/mol; Abbreviations: sq = square planar, te = tetrahedral, oct = octahedral.

	$[Ni(H_2O)_4]^{2+}$ (sq)	$[Cu(H_2O)_4]^{2+}$ (sq)	$[Zn(H_2O)_4]^{2+}$ (te)	$[Cu(H_2O)_6]^{2+}$ (oct)	$[Zn(H_2O)_6]^{2+}$ (oct)
U_{bond}	0.6254	1.5069	0.5574	0.5539	0.5469
U_{angle}	0.1112	2.3861	0.1349	0.4740	0.4196
U_{b-a}	0.0023	0.2910	0.0002	0.1013	0.0363
U_{vdW}	108.1895	79.5473	79.3527	60.5484	56.7608
U_{ele}^{perm}	-228.4007	-226.2327	-218.6961	-271.6602	-266.6295
U_{ele}^{ind}	-120.6357	-107.7846	-111.7536	-97.6227	-91.7410
$WU_{VB-bond}$	-66.1641	-17.5323	-0.3736	-36.7016	-0.5242
$WU_{VB-angle}$	38.5616	14.9848	--	36.8102	--

Table 2.6 Metal-oxygen coordination for the first solvation shell of aqueous TM ions;

MD results for present work are taken from the first peak of M-O correlation function.

	Method	1st solvation shell M-O coordination number and geometry	Reference
Cu^{2+}	MD (AMOEBA-VB)	6×2.005	Present work
	MD (REAX-FF)	$4 \times 1.94 + 2 \times 2.27$	[61]
	Neutron diffraction	6×1.97	[107]
	Neutron diffraction	5×1.96	[105]
	EXAFS	$4 \times 1.96 + 2 \times 2.60$	[108]
	EXAFS	$4 \times 2.04 + 2 \times 2.29$	[109]
	Car-Parrinello MD	5×1.96	[105]
	Car-Parrinello MD	$4 \times 2.00 + 1 \times 2.45$	[110]
Zn^{2+}	MD	6×2.055	Present work
	X-Ray diffraction	6×2.04	[100]
	B3LYP/MD	6×2.05	[101]

Chapter 3. An Angular Overlap Model in the AMOEBA

Polarizable Force Field

In this chapter, we present an AOM for Cu^{2+} ion in the AMOEBA polarizable force field. In order to demonstrate the extensibility of the AOM approach, we study the accuracy of AMOEBA-AOM for both aqueous Cu^{2+} ion and type 1 blue copper (T1Cu) proteins. Blue copper proteins, or cupredoxins, are electron transport proteins that shuttle electrons from donors to acceptors in bacteria and plants. This process takes advantage of the redox potential of Cu^{2+} and Cu^+ ions. Specifically in this study, plastocyanin (PDB: 1AG6)[111] and azurin (PDB: 1DYZ)[112] T1Cu proteins are chosen as validation targets because they are well-studied systems[113-116] with binding sites that involve most of the common ligands for Cu^{2+} ion found in biomolecules. In addition, high-resolution X-ray crystal structures are available for both of these proteins. It has been suggested that the electrostatic interactions are responsible for long-range molecular recognition of T1Cu proteins and the hydrophobic pocket near the copper binding site contributes to the precise docking of binding partners.[113] Therefore it is of interesting to apply a force field model describing both the local coordination geometry and electrostatic properties of the copper binding sites when studying these proteins.

AMOEBA-AOM force field parameters are determined against a range of gas phase QM calculations on metal complexes and validated against experimental data. In developing parameters for T1Cu proteins, small model fragments representing for protein sidechains and backbones are used in the QM-based parameterization process. Energy evaluations on gas phase metal complexes, as well as results from MD simulations of aqueous Cu^{2+} ion and T1Cu proteins are reported.

3.1 Methodology

3.1.1 AMOEBA-AOM framework

For a TM system, the total potential for the AMOEBA-AOM can be expressed as a sum of the general AMOEBA potential and the AOM energy terms specific for TM ions:

$$U_{\text{TOTAL}} = U_{\text{AMOEBA}} + U_{\text{AOM}} \tag{2.1}$$

where

$$U_{\text{AMOEBA}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{b-a}} + U_{\text{oop}} + U_{\text{torsion}} + U_{\text{vdW}} + U_{\text{ele}}^{\text{perm}} + U_{\text{ele}}^{\text{ind}} \tag{2.2}$$

The first five terms of Equation (2.2) are valence contributions representing bond stretch, angle bend, bond-angle cross-term, out-of-plane bond and torsional rotation respectively. The last three terms are nonbonded intermolecular energy terms, including the van de Waals (vdW), permanent electrostatic and induced electrostatic potentials.[58,85,86]

3.1.2 AMOEBA potentials

The details of the AMOEBA model have been previously reported. [58,85,86] For TM complexes, only the nonbonded energy terms are applied between the metal and its ligands. This is similar to the treatment of other main group cations with the exception that AOM bonding terms are used between metal ions and the atoms that are directly ligated in place of the normal vdW terms. The vdW interactions takes the form of a buffered 14-7 potential as described by Halgren[46]:

$$U_{ij}^{\text{vdW}} = \epsilon_{ij} \left(\frac{1+\delta}{\rho_{ij}+\delta} \right)^{n-m} \left(\frac{1+\gamma}{\rho_{ij}^m+\gamma} - 2 \right) \tag{2.3}$$

where $\rho_{ij} = R_{ij}/R_{ij}^0$, $n = 14$, $m = 7$, $\delta = 0.07$ and $\gamma = 0.12$. ϵ_{ij}, R_{ij}^0 and R_{ij} represent the potential energy well-depth, minimum energy distance and the separation between atoms i and j respectively. Mixing rules are applied to ϵ_{ij} and R_{ij}^0 for heterogeneous atom pairs:

$$R_{ij}^0 = \frac{(R_{ii}^0)^3 + (R_{jj}^0)^3}{(R_{ii}^0)^2 + (R_{jj}^0)^2} \tag{2.4}$$

$$\epsilon_{ij} = \frac{4\epsilon_{ii}\epsilon_{jj}}{(\epsilon_{ii}^{1/2} + \epsilon_{jj}^{1/2})^2} \tag{2.5}$$

As described below, for some ligand atom types, R_{ii}^0 is dynamically reduced via a cubic spline that is a function of ligand atom distances to the metal ion (r_{ML}):

$$R_{ii}^0 = R_{ii}^0{}' - (R_{ii}^0{}' - R_{ii}^0{}'')a_{ii}$$
$$a = c_5 r_{ML}{}^5 + c_4 r_{ML}{}^4 + c_3 r_{ML}{}^3 + c_2 r_{ML}{}^2 + c_1 r_{ML} + c_0 \tag{2.6}$$

$R_{ii}^0{}'$ is the value for minimum energy distance at metal-ligand separation beyond r_{ML}^{max}, while $R_{ii}^0{}''$ denotes the value at short range ($< r_{ML}^{min}$). This adjustment is needed to account for the reduction in atom size due to the polarization of ligand atoms towards the TM ion. The cubic spline ensures a smooth transition of R_{ii}^0 between r_{ML}^{max} and r_{ML}^{min}. The coefficients for the function are determined by imposing boundary conditions such that the dimensionless scaling factor a is 0 at r_{ML}^{max} and 1 at r_{ML}^{min}, while the first and second derivatives are 0 at r_{ML}^{max} and r_{ML}^{min}:

$$c_5 = -6/\tau$$

$$c_4 = 15(r_{ML}^{max} + r_{ML}^{min})/\tau$$

$$c_3 = -10(r_{ML}^{max2} + 4r_{ML}^{max}r_{ML}^{min} + r_{ML}^{min2})/\tau$$

$$c_2 = 30(r_{ML}^{max2}r_{ML}^{min} + r_{ML}^{max}r_{ML}^{min2})/\tau \qquad (2.7)$$

$$c_1 = -30(r_{ML}^{max2}r_{ML}^{min2})/\tau$$

$$c_0 = r_{ML}^{max3}(r_{ML}^{max2} - 5r_{ML}^{max}r_{ML}^{min} + 10r_{ML}^{min2})/\tau$$

$$\tau = \left(r_{ML}^{max} - r_{ML}^{min}\right)^5$$

The electrostatic potential consists of a permanent and an induced component. The permanent contribution is described by atom-centered monopole, dipole and quadrupole moments whose values are determined via Stone's distributed multipole analysis[87] followed by refinement against QM-derived electrostatic potentials. Polarization is handled through self-consistent induced dipole moments, with a Thole damping factor applied at short interaction distances. This mechanism has a charge smearing effect that avoids the well-known polarization catastrophe at close interatomic separations.[88]

3.1.3   AOM potentials

The complete derivations of the AOM potentials for d-row TM ion have been published elsewhere.[71,117] Here we reproduce the basic theory and its outcomes, along with modifications in the context of AMOEBA. Consider a perturbing potential v^{LF} due to the presence of ligands. Its effect on the d-orbital energies of the TM ion can be computed by first-order perturbation theory:

$$V_{ab}^{LF} = \left\langle d_a \middle| v^{LF} \middle| d_b \right\rangle \qquad (2.8)$$

The AOM makes the approximation that the ligands contribute linearly to v^{LF}, and that V_λ^{LF} is diagonal in the local frame of a ligand λ where the z-axis points away from the metal center towards the ligand atom:

$$
\begin{aligned}
\left\langle d_{\lambda,z^2} \left| v_\lambda^{LF} \right| d_{\lambda,z^2} \right\rangle &= e_\sigma = e_1 \\
\left\langle d_{\lambda,xz} \left| v_\lambda^{LF} \right| d_{\lambda,xz} \right\rangle &= e_{\pi x} = e_2 \\
\left\langle d_{\lambda,yz} \left| v_\lambda^{LF} \right| d_{\lambda,yz} \right\rangle &= e_{\pi y} = e_3 \\
\left\langle d_{\lambda,x^2-y^2} \left| v_\lambda^{LF} \right| d_{\lambda,x^2-y^2} \right\rangle &= e_{\delta x^2-y^2} = e_4 \\
\left\langle d_{\lambda,xy} \left| v_\lambda^{LF} \right| d_{\lambda,xy} \right\rangle &= e_{\delta xy} = e_5
\end{aligned}
\tag{2.9}
$$

For systems involving σ, πx and πy bondings, e_4 and e_5 can be set to zero. The orbital $|d_a\rangle$ (a = 1,2,3,4 and 5) can be expressed as a radial function multiplied by real, $l = 2$ spherical harmonics d_i. In order to develop the angular potential for the ligands, we represent the angular components of $|d_a\rangle$ as:

$$
\mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \end{pmatrix} = \begin{pmatrix} d_{z^2} \\ d_{xz} \\ d_{yz} \\ d_{x^2-y^2} \\ d_{xy} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & -1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 \\ 0 & i/\sqrt{2} & 0 & i/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 & 0 & 0 & 1/\sqrt{2} \\ -i/\sqrt{2} & 0 & 0 & 0 & i/\sqrt{2} \end{pmatrix} \begin{pmatrix} Y_{22}(\hat{r}) \\ Y_{21}(\hat{r}) \\ Y_{20}(\hat{r}) \\ Y_{2,-1}(\hat{r}) \\ Y_{2,-2}(\hat{r}) \end{pmatrix} = \mathbf{Cy} \tag{2.10}
$$

The local LF matrix must then be rotated into the global molecular frame. The spherical harmonics under a rotation R can be written as:

$$
RY_{lm}(\hat{r}) = \sum_{m'} D^l_{m'm}(\alpha\beta\gamma)Y_{lm'}(\hat{r}) \tag{2.11}
$$

where α, β and γ are Euler angles as defined in Rose[118]. For σ bonding, we can conveniently define local x-axis pointing away from the global z-axis, yielding

$$\alpha = 0, \beta = -\theta, \gamma = -\phi \tag{2.12}$$

for a ligand with polar coordinates r, θ and ϕ. In the case of non-zero πx and πy bondings, the xz-plane should be coincide with the planar ligand group. This necessitates an extra rotation through ψ, which is the angle between the new local x-axis and the one defined for σ bonding.[117] Hence:

$$\alpha = -\psi, \beta = -\theta, \gamma = -\phi \tag{2.13}$$

Rewriting Equation (2.11) in matrix form $\mathbf{D}_\lambda$, the local $\mathbf{y}_\lambda$ can be related to the global $\mathbf{y}$ by:

$$\mathbf{y}^{\mathrm{T}} = \mathbf{y}_\lambda^{\mathrm{T}} \mathbf{D}_\lambda \tag{2.14}$$

Likewise,

$$\begin{aligned} \mathbf{d}_\lambda{}^{\mathrm{T}} &= \mathbf{d}^{\mathrm{T}} \mathbf{F}_\lambda \\ \mathbf{F}_\lambda &= \mathbf{C}^* \mathbf{D}_\lambda{}^\dagger \mathbf{C}^{\mathrm{T}} \end{aligned} \tag{2.15}$$

taking advantage the fact that $\mathbf{C}$ is unitary. From there we arrive at the expression:

$$\begin{aligned} \mathbf{V}^{\mathrm{LF}} &= \sum_\lambda \mathbf{F}_\lambda \mathbf{E}_\lambda \mathbf{F}_\lambda{}^\dagger \\ E_{\lambda,ab} &= e_a \delta_{ab} \end{aligned} \tag{2.16}$$

71

If there is significant d-s hybridization, one must consider a 6×6 LF matrix involving perturbation by the (n+1)s orbital. However, Deeth *et al.*[71] has shown that this additional contribution can be simplified as

$$-\mathbf{b}\mathbf{b}^{\mathrm{T}}$$
$$b_a = \sum_\lambda F_{\lambda,a1}\sqrt{e_{\mathrm{ds}}} \tag{2.17}$$

when taking into account the fact that only $\left|d_{\lambda,z^2}\right\rangle$ can have significant overlap with $\left|d_{\lambda,s}\right\rangle$. We can then construct the overall formulation as:

$$\mathbf{V}^{\mathrm{LF}} = \mathbf{V}^\sigma + \mathbf{V}^{\pi x} + \mathbf{V}^{\pi y} - \mathbf{b}\mathbf{b}^{\mathrm{T}} \tag{2.18}$$

Diagonalizing the symmetric $\mathbf{V}^{\mathrm{LF}}$ results in energy eigenvalues w_a. Finally, combining with the occupancy of the levels (n_a), we arrive at the angular potential:

$$U_{\mathrm{AOM}}^{\mathrm{angle}} = \sum_a n_a w_a \tag{2.19}$$

In this initial iteration, a simple exponential function is used in AMOEBA-AOM for e_σ, $e_{\pi x}$, $e_{\pi y}$ and e_{ds}:

$$e_{\mathrm{AOM}} = a_{\mathrm{AOM}} r_{\mathrm{ML}}^{-6} \tag{2.20}$$

Letting m, l, $l1$ and $l2$ denote the metal, ligand atom, subsidiary atom 1 and 2 (if they exist) bonded to ligand atom respectively, we define:

$$\begin{aligned} \vec{r} &= \vec{r}_l - \vec{r}_m \\ \vec{r}_1 &= \vec{r}_{l1} - \vec{r}_l \\ \vec{r}_2 &= \vec{r}_{l2} - \vec{r}_l \end{aligned} \tag{2.21}$$

For clarity, (x,y,z) represent the components of $\vec{r}$. Then,

$$V_{ab}^{\sigma} = \sum_{ligands} G_a^{\sigma} \left[\frac{e_{\sigma}}{|\vec{r}|^4} \right] G_b^{\sigma} \tag{2.22}$$

$$b_a = \sum_{ligands} G_a^{\sigma} \left[\frac{\sqrt{e_{ds}}}{|\vec{r}|^2} \right] \tag{2.23}$$

In the case of no subsidiary atom bonded to ligand atom l,

$$V_{ab}^{\pi x} = \sum_{ligands} G_a^{\pi x} \left[\frac{e_{\pi x}}{|\vec{r}|^4} \right] \left(\frac{1}{\rho^2} \right) G_b^{\pi x}$$

$$V_{ab}^{\pi y} = \sum_{ligands} G_a^{\pi y} \left[\frac{e_{\pi y}}{|\vec{r}|^2} \right] \left(\frac{1}{\rho^2} \right) G_b^{\pi y} \tag{2.24}$$

$$\rho = \sqrt{x^2 + y^2}$$

In the case of a single subsidiary atom bonded to l,

$$V_{ab}^{\pi x} = \sum_{ligands} \left(\vec{g}_a^{\pi x} \cdot \vec{r}_1 \right) \left[\frac{e_{\pi x}}{|\vec{r}|^4} \right] \left(\frac{1}{\sigma^2} \right) \left(\vec{g}_b^{\pi x} \cdot \vec{r}_1 \right)$$

$$V_{ab}^{\pi y} = \sum_{ligands} \left(\vec{g}_a^{\pi y} \cdot \vec{r}_1 \right) \left[\frac{e_{\pi y}}{|\vec{r}|^2} \right] \left(\frac{1}{\sigma^2} \right) \left(\vec{g}_b^{\pi y} \cdot \vec{r}_1 \right) \tag{2.25}$$

$$\sigma = \sqrt{|\vec{r}_1|^2 |\vec{r}|^2 - \left(\vec{r}_1 \cdot \vec{r} \right)^2}$$

Finally, in the case of two subsidiary atoms bonded to l,

$$V_{ab}^{\pi x} = \sum_{ligands} \left(\vec{h}_a^{\pi x} \cdot (\vec{r}_1 \times \vec{r}_2) \right) \left[\frac{e_{\pi x}}{|\vec{r}|^2} \right] \left(\frac{1}{\tau^2} \right) \left(\vec{h}_b^{\pi x} \cdot (\vec{r}_1 \times \vec{r}_2) \right)$$

$$V_{ab}^{\pi x} = \sum_{ligands} \left(\vec{h}_a^{\pi y} \cdot (\vec{r}_1 \times \vec{r}_2) \right) \left[\frac{e_{\pi y}}{|\vec{r}|^4} \right] \left(\frac{1}{\tau^2} \right) \left(\vec{h}_b^{\pi y} \cdot (\vec{r}_1 \times \vec{r}_2) \right)$$

$$\vec{h}_a^{\pi x} = \vec{g}_a^{\pi y}$$

$$\vec{h}_a^{\pi y} = -\vec{g}_a^{\pi x}$$ (2.26)

$$\tau = \sqrt{|\vec{r}_1 \times \vec{r}_2|^2 |\vec{r}|^2 - \left((\vec{r}_1 \times \vec{r}_2) \cdot \vec{r} \right)^2}$$

Expressions for G_a^σ, $G_a^{\pi x}$, $G_a^{\pi y}$, $\vec{g}_a^{\pi x}$ and $\vec{g}_a^{\pi y}$ are reproduced in Table 3.1.

Our AMOEBA-AOM differs from other implementations of similar models in MM force fields[73] in that the classical electrostatic model is applied consistently to both the TM and its ligands. This setup allows the study of ligand exchanges since the AOM energy terms drops off rapidly with increasing metal-ligand separation but electrostatic contributions remains significant at distances beyond ligation shell. It should be noted that retaining the electrostatic model affects the parameterization of a_{AOM} and therefore our parameters are not directly comparable with previously reported values.

The metal-ligand bonding interaction is described by a Morse potential:

$$U_{AOM}^{bond} = D(1 - e^{-a_{Morse}(r_{ML} - r_{ML,0})^2}) - D$$ (2.27)

where D, a_{Morse} and $r_{ML,0}$ controls the bond strength, width of the potential well and the minimum energy distance respectively.

3.2 Parameterization and Validation

The AMOEBA-AOM parameters were determined via methods similar to previously published parameterization routines for the AMOEBA-VB model.[119] The

general strategy was to fit the MM results of energy evaluations and geometry optimizations to those obtained via QM calculations on gas phase TM-complexes under a variety of different conditions. The AOM parameters were determined after the AMOEBA parameters had been finalized following the usual protocol.[86] The goal of the parameterization process is to obtain a single set of AOM parameters that best reproduces the QM results for all test routines. Finally, analyses based on MD simulation results were validated against available experimental and computational data.

All *ab initio* calculations were carried out with the Gaussian 09[93] software. Geometry optimization of aqua Cu^{2+} complexes were performed at the B3LYP/6-311G(d,p) [21,23,82] level of theory. Single-point energies were evaluated using the MP2 electron correlation method[120], with the aug-cc-pVTZ[94] basis set for main group atoms and cc-pVTZ[95] on Cu^{2+} ion. A Fermi-broadening SCF technique[96] was used to improve convergence stability, and a relatively stringent SCF convergence criterion of 10^{-9} a.u. was imposed. In the case of model complexes for the Cu^{2+} binding sites in T1Cu proteins, B2PLYP-D/cc-pVDZ[24,25] and MP2/cc-pVDZ were utilized for geometry optimizations and for single-point energy calculations respectively. Ligand internal coordinates were frozen during optimization calculations to increase computational efficiency. The AMOEBA-AOM energy terms and their corresponding analytical derivatives were implemented in the TINKER[58] MM package.

3.2.1   *Gas phase calculations on aqua Cu^{2+} complexes*

The AMOEBA water parameters have been reported previously[85] and were unmodified for use with AMOEBA-AOM. QM geometry optimizations were performed

on gas phase tetra-aqua and hexa-aqua Cu^{2+} complexes under angular constraints to yield idealized square-planar, tetrahedral and octahedral structures. The following procedures were used to compare MM and QM calculations performed on geometric variants generated from these optimized complexes:

1 Copper-water bonding potential curves were produced for square-planar $[Cu(H_2O)_4]^{2+}$ and octahedral $[Cu(H_2O)_6]^{2+}$ by performing single-point energy evaluations at varying copper-oxygen distances for a single water molecule. Axial and in-plane water molecules in $[Cu(H_2O)_6]^{2+}$ are monitored separately to illustrate the effect of the Jahn-Teller distortion. Zero energy is taken to be the potential of complex with copper-oxygen distance at 5Å.

2 The potential energy difference between square-planar and tetrahedral $[Cu(H_2O)_4]^{2+}$ are plotted as a function of copper-oxygen separations, with water-water interactions removed. This gives a direct measurement to the LF effect since it is known that 4-coordinated Cu^{2+} complexes do not adopt the tetrahedral geometry for small ligands which minimizes water-water repulsion.[14,43,77]

3 One hundred complex structures were generated by introducing small geometric perturbations to the optimized square-planar $[Cu(H_2O)_4]^{2+}$ and octahedral $[Cu(H_2O)_6]^{2+}$. This process involves randomly perturbing the copper-oxygen distance by a maximum of ± 0.2Å deviation from optimal value, as well as rotating each of the water molecules around the copper-oxygen vector and two orthogonal axes between 0 and 10 degrees. Structures containing a water-water separation less than 2.5Å were discarded. MM computed energies for these complexes were compared to the results

obtained from QM to investigate whether MM models can reproduce the QM energy surface near the optimum geometry. Structures with QM energies more than 15 kcal/mol higher than that of the idealized geometry were discarded since these high-energy structures are not readily accessible during routine MD simulations.

Procedural diagrams for routines described above are available in Figure 3.1.

3.2.2 Gas phase calculations on model complexes for Cu^{2+} binding sites in T1Cu proteins

The Cu^{2+} binding site of 1AG6 plastocyanin consists of two histidine, one cysteine and one methionine residue.[111] In addition to these ligands, the copper ion is coordinated by an extra backbone carbonyl oxygen in the structure of 1DYZ azurin.[112] The structures of the Cu^{2+} binding sites are visualized in Figure 3.2. For gas phase calculations performed during the AOM parameterization process, complete amino acid residues were substituted by small model compounds, which were chosen to maintain similar ligand properties. The identities of the corresponding model fragments can be found in Table 3.2. For the sake of brevity, the model complexes for the 1AG6 and 1DYZ Cu^{2+} binding sites will be denoted by T1Cu1 and T1Cu2 respectively in the following discussions. The AMOEBA parameters for the ligands were obtained via the published protocol and their values can be found in the Appendix B. Similar to the procedure used for water molecules, the AOM parameters were obtained by fitting results from a series of MM computations to those obtained from QM:

1 Geometry optimizations were carried out for T1Cu1 and T1Cu2 using both QM and MM. The ligation geometries of the optimized structures were compared.

2   QM binding energies are computed by performing counterpoise-corrected MP2 calculations on B2PLYP-D optimized structures with the ligand and the rest of the complex in two different fragments. The data is then compared to MM interaction energies that are calculated by subtracting the potential energies of the individual ligand and the remaining molecules from the overall complex energy.

3   Random complex structures were generated for T1Cu1 and T1Cu2 following a similar protocol to that applied to aqua Cu^{2+} complexes. The ligand molecules are rotated from the QM optimized geometry by a maximum of 15 degrees with respect to metal-ligand vector, defined by the Cu^{2+} ion and atom directly ligated to the metal, and two orthogonal axes. A minimum ligand-ligand contact distance of 2.5Å is maintained. Sets of one hundred structures were generated for each ligand and only a single ligand is perturbed within each set. Geometries with *ab initio* energy higher than 5kcal/mol from those of the QM optimized complexes were discarded when comparing QM and MM potentials.

3.2.3   *Aqueous Cu^{2+} simulations*

Canonical ensemble MD simulations were performed on a single Cu^{2+} ion solvated in a 18.6215Å cubic water box. Period boundary condition was enforced and particle-mesh Ewald summation was applied to long-range electrostatic interactions.[98,99] Self-consistent dipole polarization was converged to 0.01 Debye root-mean-squared (RMS) change in atomic induced dipole moments. Multiple 10ns trajectories generated with 1fs time-step were collected at temperatures of 298K, 320K, 350K and 380K. The correlation function, solvation shell properties, coordination numbers and water residence

times were calculated from each of the trajectories and compared against previous published data. The first 100ps of the trajectories were discarded to allow as system equilibration.

3.2.4   T1Cu protein simulations

MD simulations were carried out at 298K in the canonical ensemble for 1AG6 and 1DYZ proteins. The available AMOEBA protein parameters (parameter file: amoebabio09.prm) were used[53] while the AMOEBA-AOM parameters derived from T1Cu1 and T1Cu2 were applied to the appropriate residues. Water molecules external to the proteins were first removed from the X-ray structures. Hydrogen atoms were then added, with positions determined from heavy-atom bonding geometries. The protonation state of histidine residues were assigned by analyzing the local hydrogen-bonding network.[121] Additionally, unresolved atoms were filled in manually to construct a full side chain for GLU19 of 1DYZ. The protein structures were solvated in water inside a 98.6726Å truncated octahedron. Before simulations were conducted, the water molecules coordinates were minimized to 3 kcal/mol RMS change in potential energy gradient, followed by minimization on the entire system to 2 kcal/mol. Settings for dipole polarization and long-range electrostatics were identical to those used in the simulations for aqueous Cu^{2+} and periodic boundary conditions were applied. A total of 500ps of MD trajectory was collected for each protein. The geometries of the Cu^{2+} binding sites were compared against previously published experimental and computational studies.

3.3 Results and Discussions

3.3.1   AMOEBA-AOM parameters

The AMOEBA parameters for Cu^{2+} ion are identical to those used in our previous AMOEBA-VB study.[119] The AOM parameters for water, T1Cu1 and T1Cu2 ligands are presented in Table 3.3. A number of constraints on the values of the AOM parameters are applied during the parameterization process. First, e_σ should be the largest contribution to the AOM matrix, since it represents the principle LF. Secondly, the $e_{\pi x}$ term is zero for ligand atoms with two bonded subsidiary atoms as the local y-axis is taken to be perpendicular to the ligand plane. Finally, $e_{\pi x}$ and $e_{\pi y}$ have equal values in case of ligand atoms with a single bonded subsidiary atom because the contributions from ligand orbitals should be cylindrical. A common set of AOM parameters were used in all the calculations presented here.

3.3.2 Gas phase calculations on aqua Cu^{2+} complexes

The bonding potentials of water molecules for square-planar $[Cu(H_2O)_4]^{2+}$ and octahedral $[Cu(H_2O)_6]^{2+}$ are plotted in Figure 3.3. Both AMOEBA and AMOEBA-AOM can reproduce the QM minimum energy distance for $[Cu(H_2O)_4]^{2+}$ but AMOEBA underestimates the strength of interaction by 4.5 kcal/mol whereas AMOEBA-AOM (-39.8 kcal/mol) is in better agreement with QM results (-40.3 kcal/mol). For $[Cu(H_2O)_6]^{2+}$, data from AMOEBA and AMOEBA-AOM are comparable for in-plane water molecules. However, AMOEBA is not able to capture the distortion of axial water molecules while AMOEBA-AOM can reasonably describe the structural extent of the Jahn-Teller distortion. The QM-derived bonding distance for an axial water is 2.3Å, compared to 2.1Å and 2.2Å for AMOEBA and AMOEBA-AOM respectively. In addition, AMOEBA-AOM (-20.5 kcal/mol) generates a binding energy closer to that of QM (-18.0 kcal/mol) than AMOEBA (-24.2 kcal/mol).

Figure 3.4 shows the potential energy differences between square-planar and tetrahedral $[Cu(H_2O)_4]^{2+}$ complexes at varying copper-oxygen distances. It is evident that without the AOM terms, AMOEBA produces the wrong geometric preference for $[Cu(H_2O)_4]^{2+}$. The AMOEBA-AOM model correctly prefers the square-planar geometry and the computed energy difference is in good agreement with the QM results.

Figure 3.5 compares the QM and MM computed energy surfaces near the optimized square-planar $[Cu(H_2O)_4]^{2+}$ and octahedral $[Cu(H_2O)_6]^{2+}$. All the values presented are relative to the potential of the idealized structures. The addition of the AOM term significantly reduces the RMS deviation from *ab initio* results for $[Cu(H_2O)_4]^{2+}$ (0.72 to 0.38 kcal/mol). Interestingly, the performance of AMOEBA and AMOEBA-AOM are comparable for $[Cu(H_2O)_6]^{2+}$. We attribute this to the fact that the ligands are in close contact in a 6-coordinated complex and therefore ligand-ligand interaction plays a predominant role in determining the energy surface. The small perturbations introduced to metal-ligand distances were not significant enough to demonstrate the effect of the Jahn-Teller distortion.

3.3.3 *Aqueous Cu^{2+} ion simulations*

The copper-oxygen pairwise correlation function and radial distribution are computed from MD simulations performed at 298K, 320K, 350K and 380K (Figure 3.6). An occupancy of 5-6 in the first solvation shell has been previously purposed for aqueous Cu^{2+} in the literature.[97,102-106] It has also been suggested that performing simulation at elevated temperature can result in a transition of the coordination number from 5 to 6.[105] In this study, we are unable to find evidence for 5-coordinate solvation. The radial

distribution plot suggests a 6-coordinate first solvation shell at all simulation temperatures. This result echoes the observations we made in our previous study on aqueous Cu^{2+} ion using the AMOEBA-VB model.[119] The lower peak value of the correlation function at the higher temperatures indicates a less structured solvation shell. In addition, we are again unable to observe the "dual-peak" character previously obtained from simulation carried out with ReaxFF model.[61] Comparisons of the coordination geometries taken from present and prior reports can be found in Table 3.4.

The residence times of water molecules in in the first solvation shell are computed (Table 3.5) by counting the number of continuous frames a particular water oxygen atom spends within 3Å to the Cu^{2+} ion. This cutoff distance is determined by inspecting the mid-point separation of first and second solvation shell as indicated in the pairwise correlation function (Figure 3.6). Transient water molecules with less than a 1ps presence and significantly elongated average Cu-O distances (> 2.8Å) are excluded to avoid skewing the statistics. Using the AMOEBA-AOM, we obtained an average residence time of 4ns at room temperature, which agrees with the most recent NMR-based experimental value of 5ns. Older values ranging from 20ns to 0.4μs have been reported but are subject to considerable uncertainty.[97,122,123] The residence time is much shorter than previously reported room-temperature experimental values for other third row TM ions such as Ni^{2+} (37μs) and Fe^{2+} (0.3μs) but longer than Zn^{2+} (0.1-5ns). [97,124] As expected, we observed a shortening of residence time with increasing simulation temperature.

3.3.4   *Gas phase calculations on T1Cu1 and T1Cu2*

Table 3.6 summarizes the geometries of optimized T1Cu1 and T1Cu2 structures using QM and MM. A visual overlap of optimization results from QM and AMOEBA-AOM is presented in Figure 3.7. In general, the results computed with the AMOEBA-AOM agree reasonably well with QM structures. The AMOEBA-AOM yields significantly better angular geometry than AMOEBA, which is expected since standard AMOEBA lacks any explicit description of electronic LF effects. It is of interest to point out some discrepancies between the AMOEBA-AOM and QM structures. The geometry obtained from B2LYP-D optimization shows significant elongation in copper-dimethyl sulfide distance in T1Cu2 compared to T1Cu1. This property is not well described by the AMOEBA-AOM in its current version. A possible explanation is that some of the AOM parameters may be better described by a different function of the metal-ligand distance. The parameters reported were fitted to produce a binding distance of approximately 2.8Å, which is a commonly observed value for copper-methionine ligation in T1Cu proteins.[125] Furthermore, there is significant deviation from the QM value of the dimethyl sulfide – metal – imidazole 2 angle in T1Cu1. This discrepancy may be coupled to the difference in binding distances for the dimethyl sulfide ligand.

The binding energies for T1Cu1 and T1Cu2 ligands computed by QM and MM can be found in Table 3.7. In this context, the AMOEBA-AOM is an improvement over AMOEBA for both the imidazole and acetamide ligands. AMOEBA performs remarkably well for ethyl thiolate, considering the close proximity between two highly charged atoms. However, the AMOEBA-AOM has difficulty in treating some sulfur ligands, especially the dimethyl sulfide ligand in T1Cu2. Nevertheless, the overall energy values are reasonable for this initial implementation of the AMOEBA-AOM. Further

refinement of parameters against a larger set of training complexes should improve the results.

Comparisons of QM and MM potentials of random T1Cu1 and T1Cu2 structures are shown in Figure 3.8. The addition of the AOM energy term dramatically improves the overall correlation between QM and MM computed potentials. There is a 73% and 64% reduction in RMS error for T1Cu1 and T1Cu2 complexes respectively. It can be observed that sets of structures with perturbations to sulfur-type ligands result in the largest deviations of the AMOEBA-AOM energies from *ab initio* potentials.

3.3.5   T1Cu proteins simulations

The RMS distances from the initial PDB experimental coordinates for backbone alpha-carbon atoms, as well as copper-binding side chain and carbonyl atoms, are plotted in Figure 3.9 and 3.10 respectively. The RMS superposition for the backbone suggests that the protein maintains the same general fold as the X-ray structure throughout the course of the simulation. The ensemble average geometries of Cu^{2+} binding sites (Table 3.8) are computed based on atomic coordinates, excluding the first 50ps of each trajectory. In general, the ligation geometry of Cu^{2+} binding sites obtained from MD simulation agrees reasonably well with the X-ray crystal structures. The main difference between simulated and experimental structure is again the methionine binding distance in 1DYZ azurin. The computed average Cu^{2+}-MET121 distance is about 0.4Å too short, similar to the observations we made for T1Cu2 model complex. This discrepancy has also been found in other computational studies on azurin.[114,125]Overall, the performance

of the AMOEBA-AOM on plastocyanin and azurin is comparable to previously purposed MM models.[114,115,125]

3.4 Conclusions

The AMOEBA-AOM is an extensible polarizable force field for TM ions that is suitable for studying a variety of TM systems. Its principle advantages over most other AOM-based MM models for TM ion is in the consistent treatment of electrostatics at all distances and explicit description of polarization, which in turn enables the study of ligand association/dissociation and other dynamic events. We have demonstrated that the AMOEBA-AOM provides excellent agreement with QM for a wide range of calculations on aqua Cu^{2+} complexes. It also automatically handles the Jahn-Teller distortion for hexa-aqua Cu^{2+} complexes. The computed aqueous Cu^{2+} ligation geometry and water residence time in the first solvation shell are in line with published experimental results. In addition, we have provided evidence for parameter transferability in the context of the T1Cu proteins, yielding reasonable results when compared to gas-phase QM calculations on model complexes and X-ray crystallographic ligation data for complete proteins. Finally, the AMOEBA-AOM is much more efficient than semi-empirical or hybrid QM methods, allowing us to perform MD simulations on T1Cu systems investigated in this report that consisting upward of 48,000 atoms.

3.5 Figures
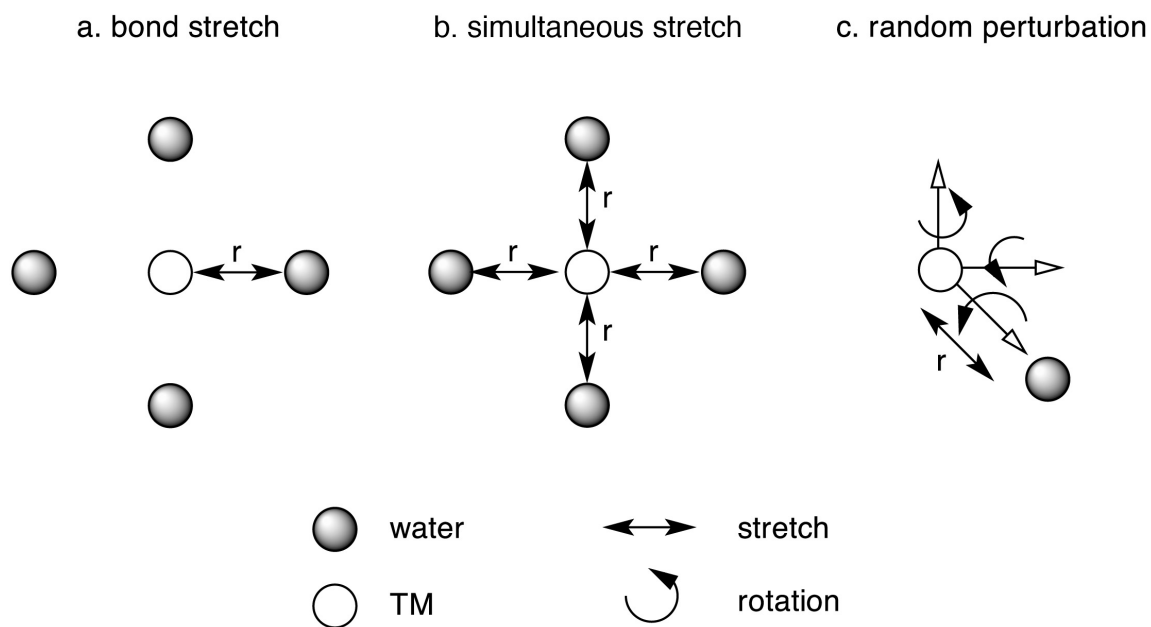
Figure 3.1 Routines for generating structural variants from QM-optimized aqua Cu^{2+} complexes for use in the AMOEBA-AOM parameterization process. (a) A single copper-water distance is varied while other ligands retain their optimized coordinates. (b) All copper-water distances are changed simultaneously with each ligand equidistant from the copper ion. (c) Random perturbations are introduced by varying copper-water distances as well as by rotating the ligands with respect to the copper-oxygen vector and two axes orthogonal to the vector.

1AG6   1DYZ
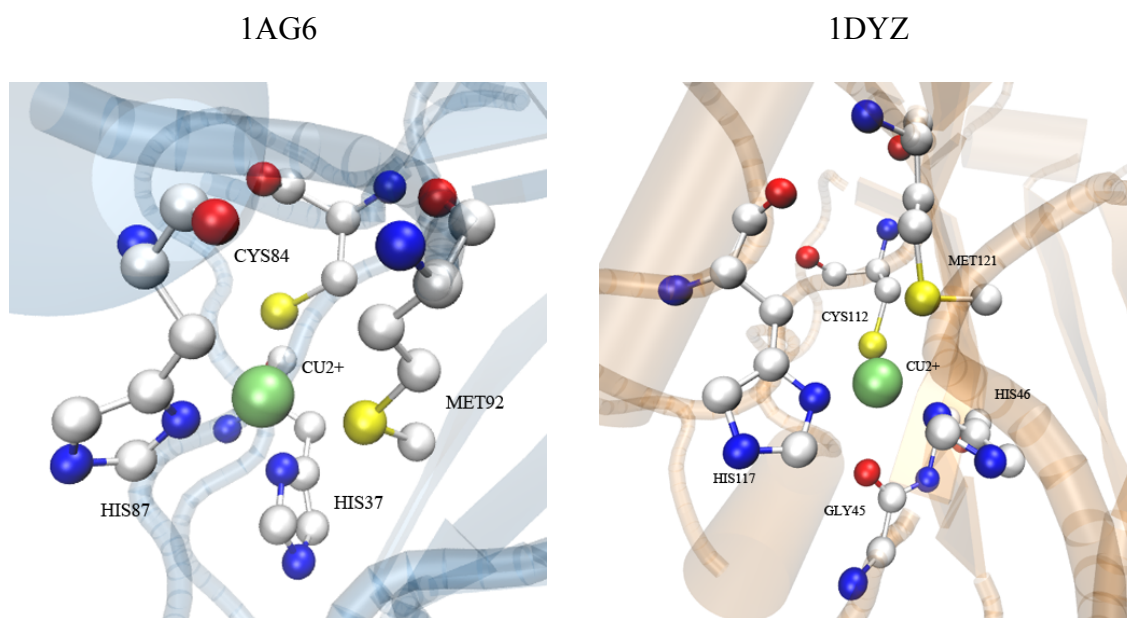
Figure 3.2 Visual representations of Cu^{2+} binding sites in X-ray structures of 1AG6 and 1DYZ. Colors: Cu^{2+} = lime green, oxygen = red, nitrogen = blue, sulfur = yellow, carbon = white.
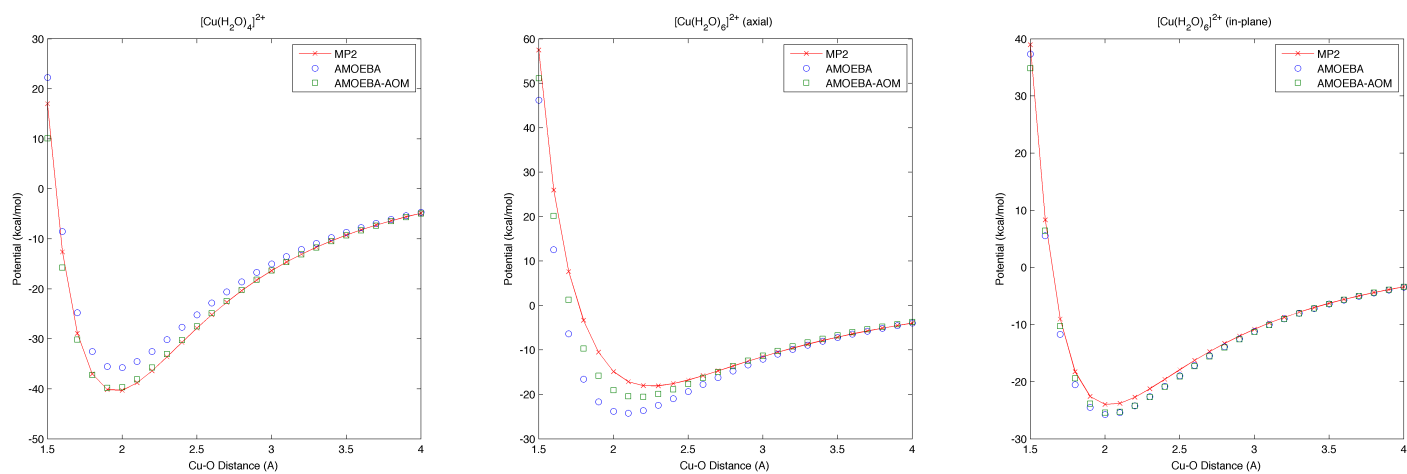
Figure 3.3 Bonding potential curve of water molecule generated using QM and MM methods. Zero bonding potential energy is taken as the potential of the complex with a water molecule at 5Å.
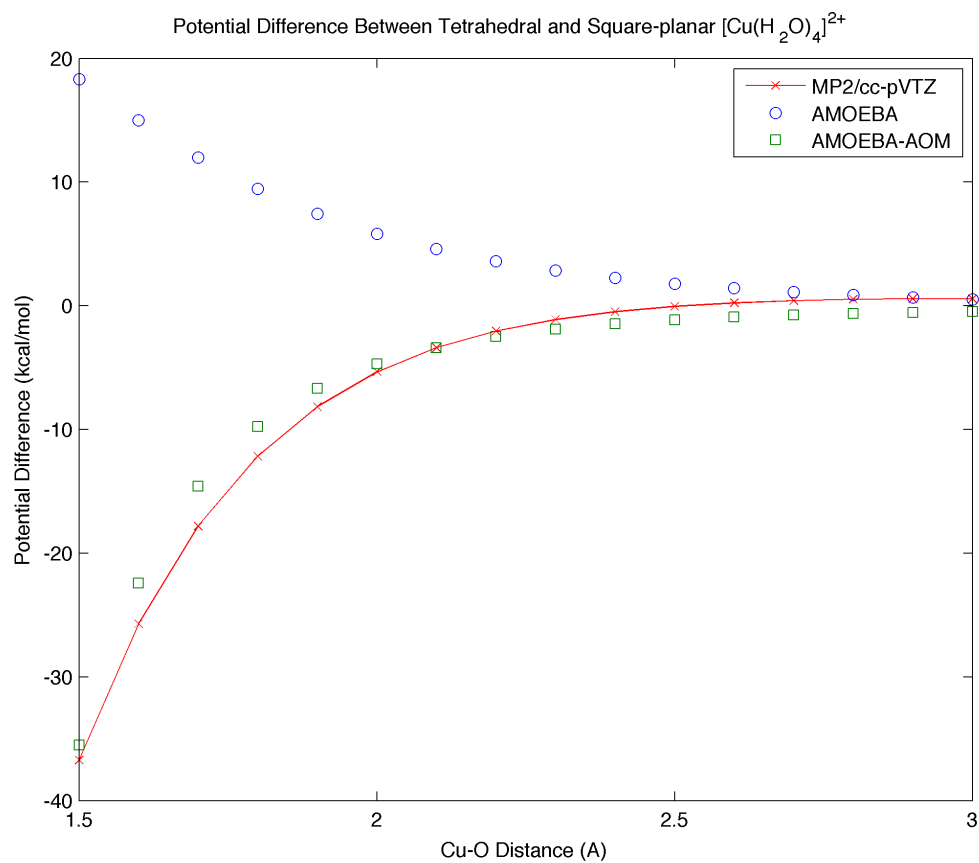
Figure 3.4 Potential energy difference between square-planar and tetrahedral tetra-aqua Cu^{2+} complexes with the water-water interaction removed. Negative values indicate that the square-planar structure is lower in potential energy than the tetrahedral geometry.
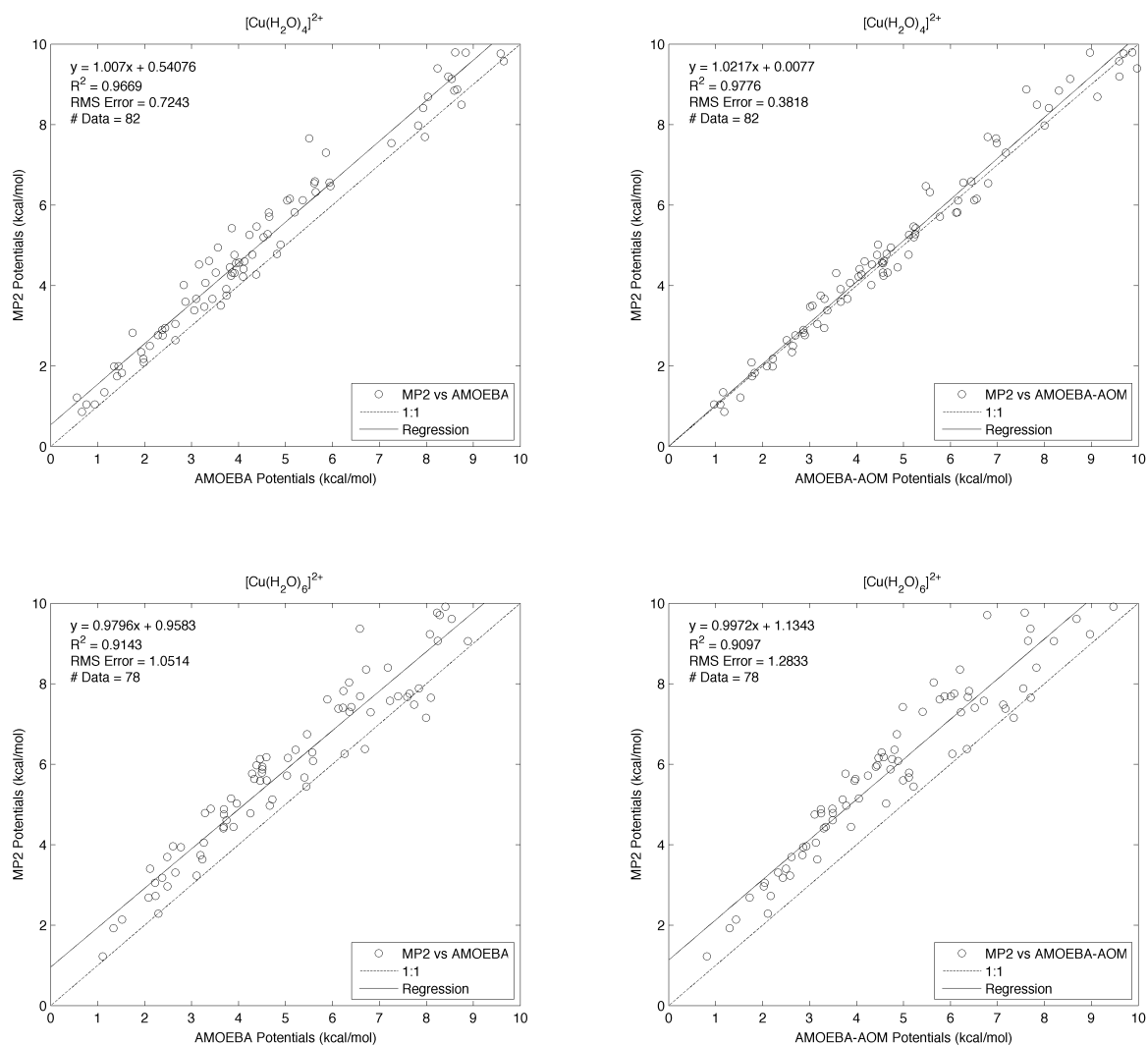
Figure 3.5 Comparisons between QM and MM potentials of random aqua Cu^{2+} complexes generated by perturbing the QM-optimized structure.
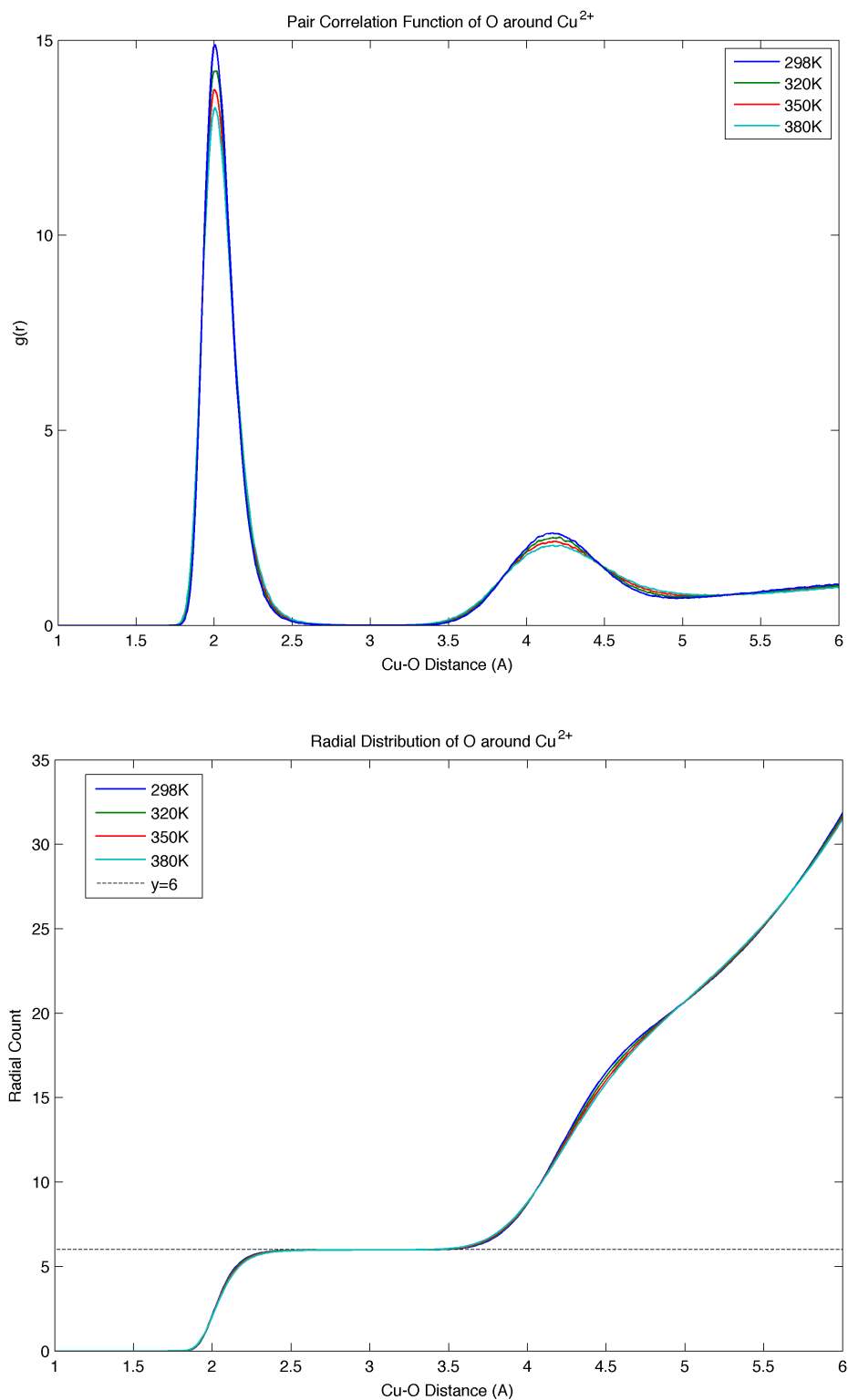
Figure 3.6 Copper-oxygen radial pair-wise correlation (above) and distribution function (below) computed for MD trajectories at various simulation temperatures.

T1Cu1 T1Cu2
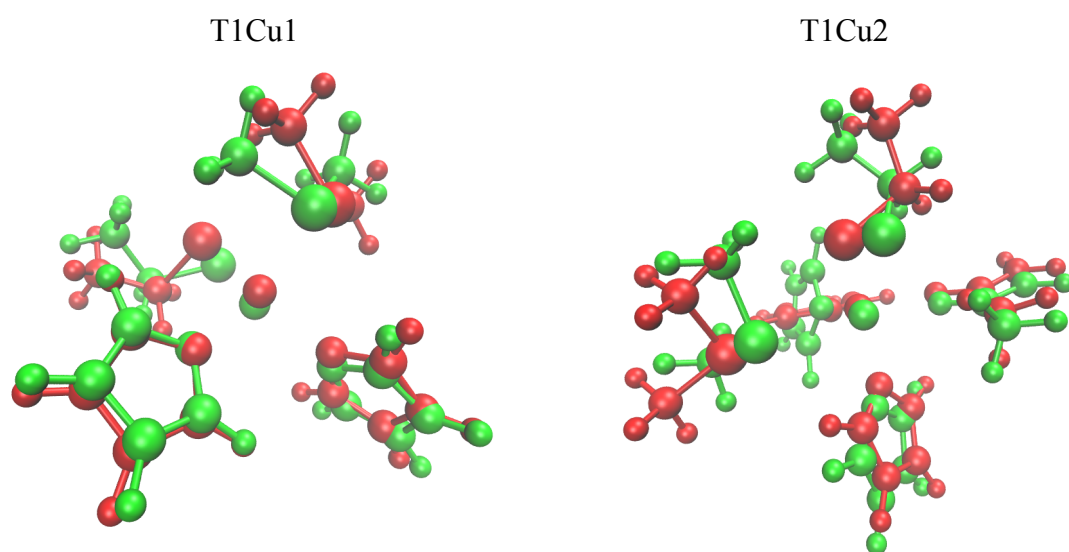
Figure 3.7 Structures of T1Cu1 and T1Cu2 optimized using B2PLYP-D/cc-pVDZ and AMOEBA-AOM. Colors: QM = red, AMOEBA-AOM = green.
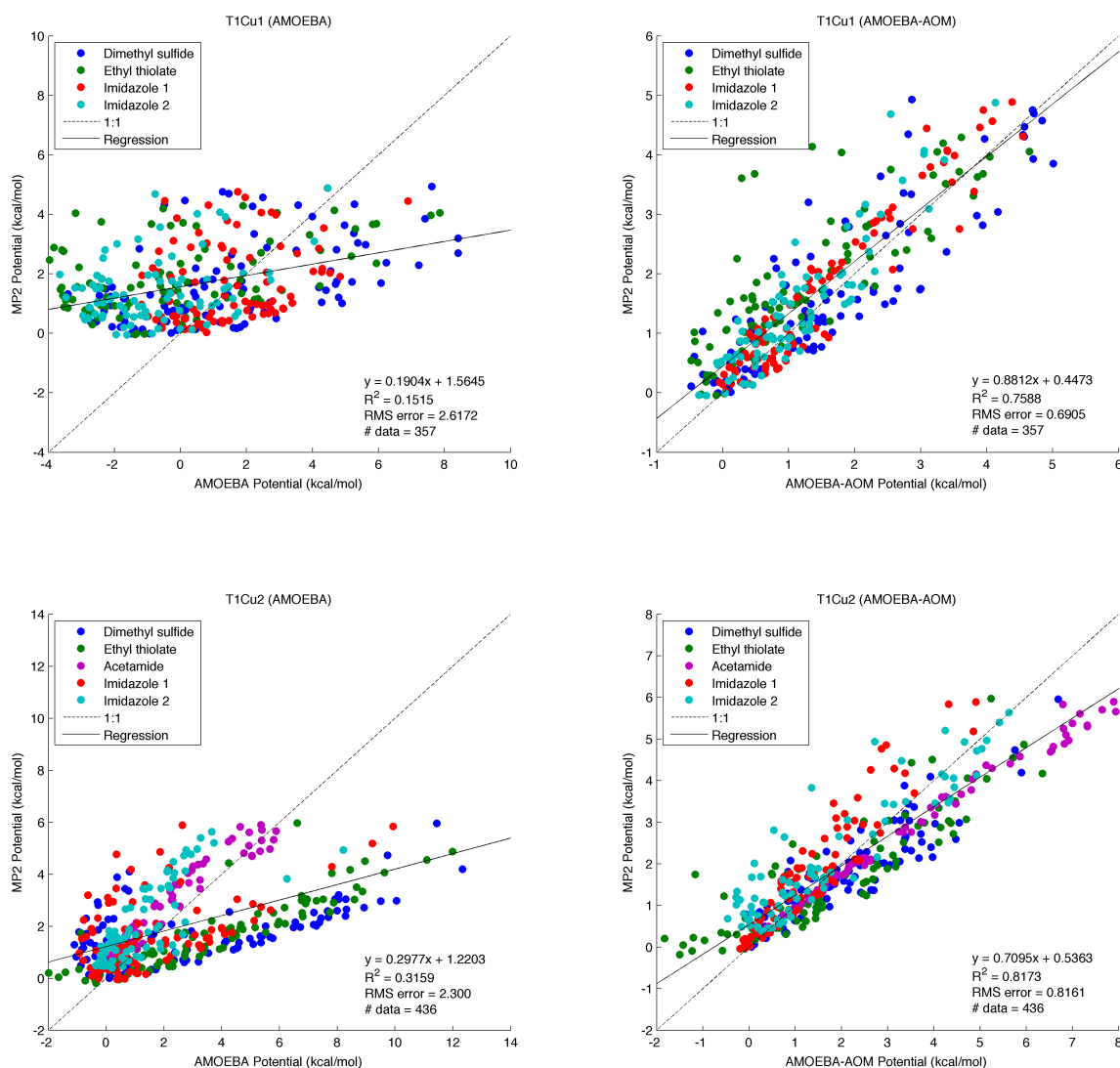
Figure 3.8 Comparison of QM and MM potentials of random T1Cu1 and T1Cu2 complexes. Results obtained from AMOEBA are plotted on the left column and those computed with the AOM energy terms are on the right. Data point colors represent different sets of structures generated by perturbing a particular type of ligand. Plots of individual ligands can be found in the Appendix C.
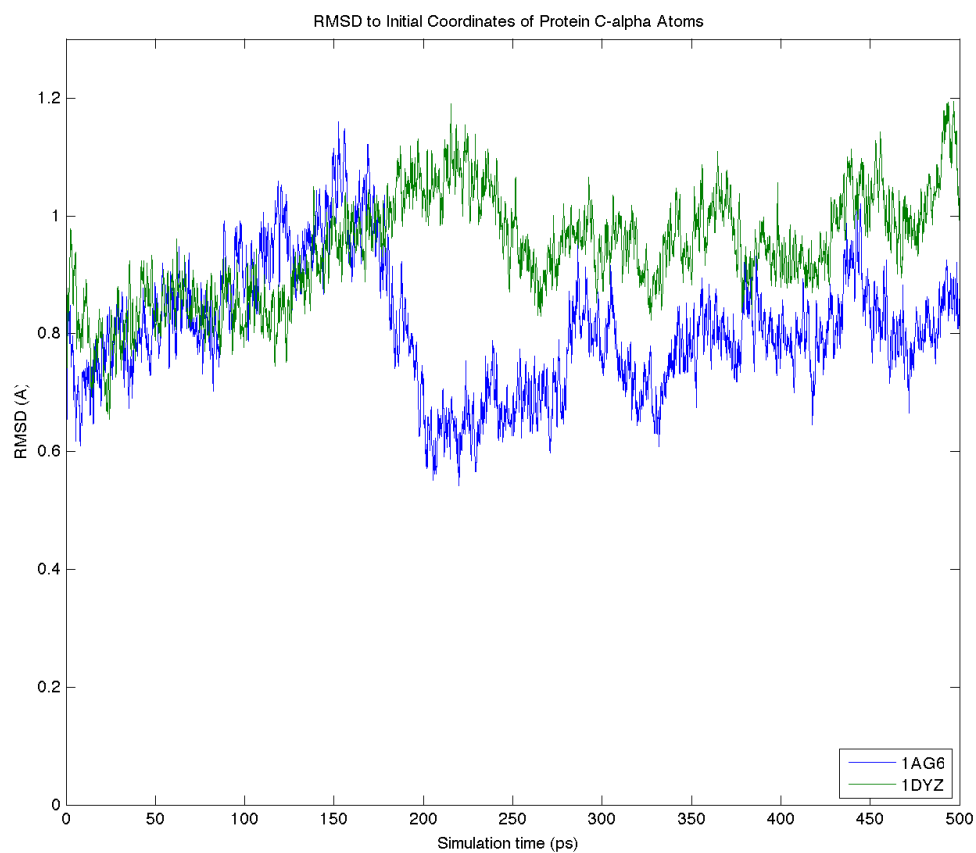
Figure 3.9 Time evolution of the RMS distance to the initial protein alpha carbon atoms.
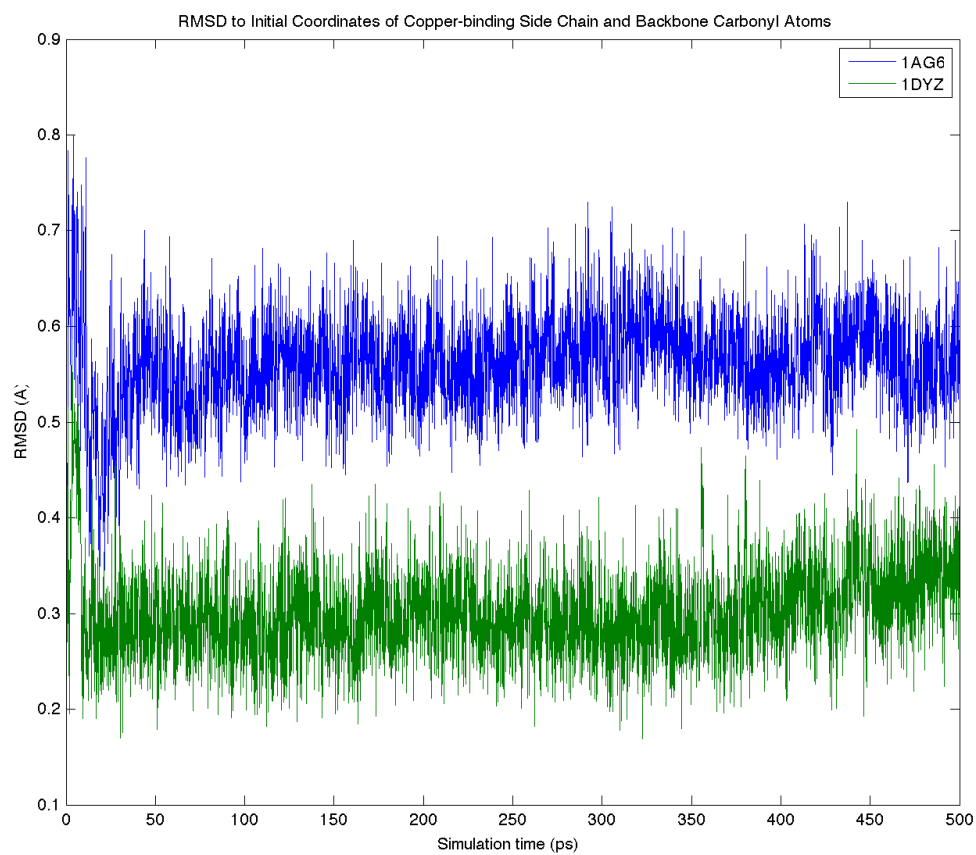
Figure 3.10 Time evolution of the RMS distance to the initial protein structure after superposition of copper-binding side chain and backbone carbonyl atoms.

3.6 Tables

Table 3.1 Expressions for AOM terms G_a^σ, $G_a^{\pi x}$, $G_a^{\pi y}$, $\vec{g}_a^{\pi x}$ and $\vec{g}_a^{\pi y}$. $(x,\ y,\ z)$ are components of the metal-ligand vector.

a	G_a^σ	$G_a^{\pi x}$	$G_a^{\pi y}$
1	$\frac{1}{2}\left(2z^2 - x^2 - y^2\right)$	$-\sqrt{3}\left(x^2 + y^2\right)z$	0
2	$\sqrt{3}xz$	$x\left(z^2 - x^2 - y^2\right)$	$-yz$
3	$\sqrt{3}yz$	$y\left(z^2 - x^2 - y^2\right)$	xz
4	$\frac{1}{2}\sqrt{3}\left(x^2 - y^2\right)$	$z\left(x^2 - y^2\right)$	$-2xy$
5	$\sqrt{3}xy$	$2xyz$	$x^2 - y^2$

a	$\vec{g}_{a,x}^{\pi x}$	$\vec{g}_{a,y}^{\pi x}$	$\vec{g}_{a,z}^{\pi x}$
1	$-\sqrt{3}xz^2$	$-\sqrt{3}yz^2$	$-\sqrt{3}\left(x^2 + y^2\right)z$
2	$z\left(z^2 - x^2 + y^2\right)$	$-2xy$	$x\left(x^2 + y^2 - z^2\right)$
3	$-2xyz$	$z\left(x^2 - y^2 + z^2\right)$	$y\left(x^2 + y^2 - z^2\right)$
4	$x\left(2y^2 + z^2\right)$	$-y\left(2x^2 + z^2\right)$	$-z\left(x^2 - y^2\right)$
5	$y\left(z^2 - x^2 + y^2\right)$	$x\left(x^2 - y^2 + z^2\right)$	$-2xyz$

a	$\vec{g}_{a,x}^{\pi y}$	$\vec{g}_{a,y}^{\pi y}$	$\vec{g}_{a,z}^{\pi y}$
1	$-\sqrt{3}yz$	$\sqrt{3}xz$	0
2	$-xy$	$x^2 - z^2$	yz
3	$z^2 - y^2$	xy	$-xz$
4	$-yz$	$-xz$	$2xy$
5	xz	$-yz$	$y^2 - x^2$

Table 3.2 Corresponding model fragments used in QM gas phase calculations to model copper binding sites of T1Cu proteins.
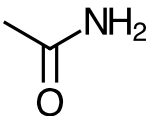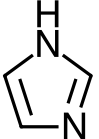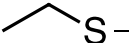
Binding site ligands	Model compound
Backbone carbonyl	![acetamide structure] (acetamide)
Histidine side chain	![imidazole structure] (imidazole)
Methionine side chain	![dimethyl sulfide structure] (dimethyl sulfide)
Cysteine side chain (deprotonated)	![ethyl thiolate structure] (ethyl thiolate)

Table 3.3 The AOM parameters for water, T1Cu1 and T1Cu2 ligands defined by the bolded atoms. See Equation (2.6), (2.20) and (2.27) for variable definitions. Ligands with the same value of $R_{ii}^{0}{}'$ and $R_{ii}^{0}{}''$ indicates that vdW scaling is not applied. r_{ML}^{min} and r_{ML}^{max} are set at 4.5Å and 6Å respectively for all ligands.
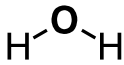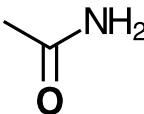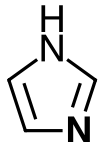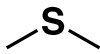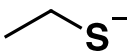
Ligand	H–O–H	(NH₂ amide)	(imidazole)	(S methyl)	(S⁻ thiolate)
a_σ	110	13811	494	6269	5360
a_{ds}	90	2170	4	793	1664
$a_{\pi x}$	0	1200	0	0	265
$a_{\pi y}$	5	1200	106	132	265
D	1.160	130.0	10.00	18.00	1.00
a_{Morse}	1.810	3.950	1.750	2.900	1.000
$r_{ML,0}$	2.835	2.200	2.500	2.800	4.000
$R_{ii}^{0}{}'$	1.703	1.650	1.855	2.000	2.175
$R_{ii}^{0}{}''$	1.703	1.650	1.705	1.950	2.000

Table 3.4 The 1st solvation shell coordination geometry of aqueous Cu^{2+} ion. Value for the present work is taken from the first peak of the copper-oxygen pairwise correlation function generated at 298K.

Method	1st solvation shell M-O coordination number and geometry	Reference
MD (AMOEBA-AOM)	6×2.005	Present work
MD (AMOEBA-VB)	6×2.005	[119]
MD (REAX-FF)	$4 \times 1.94 + 2 \times 2.27$	[61]
Neutron diffraction	6×1.97	[107]
Neutron diffraction	5×1.96	[105]
EXAFS	$4 \times 1.96 + 2 \times 2.60$	[108]
EXAFS	$4 \times 2.04 + 2 \times 2.29$	[109]
Car-Parrinello MD	5×1.96	[105]
Car-Parrinello MD	$4 \times 2.00 + 1 \times 2.45$	[110]

Table 3.5 The residence times of water molecules in the first solvation shell computed by counting the number of frames a water oxygen atom is spent within 3Å to the Cu^{2+} ion.

Temperature	Residence Time (ns)
298K	4.0
320K	2.6
350K	1.4
380K	1.2

Table 3.6 Geometries of optimized T1Cu1 and T1Cu2 complexes using DFT, AMOEBA

and AMOEBA-AOM methods.

	T1Cu1			T1Cu2		
	B2LYP-D	AMOEBA-AOM	AMOEBA	B2PLYP-D	AMOEBA-AOM	AMOEBA
Metal-ligand bond length (Å)						
Ethyl thiolate	2.20	2.08	2.33	2.12	2.24	2.35
Dimethyl sulfide	2.41	2.84	2.41	3.50	2.78	4.13
Imidazole 1	2.07	2.32	1.98	2.00	2.36	2.00
Imidazole 2	2.20	2.36	1.99	2.02	2.33	2.00
Acetamide	-	-	-	2.38	2.49	1.92
Ligand-metal-ligand angle (º)						
Ethyl thiolate – Dimethyl sulfide	94.38	105.75	107.99	79.02	90.77	69.42
Ethyl thiolate – Imidazole 1	148.41	147.40	112.46	123.78	123.38	113.68
Ethyl thiolate – Imidazole 2	99.54	118.94	115.84	132.95	145.13	113.37
Dimethyl sulfide – Imidazole 1	90.11	87.37	103.26	91.48	92.34	80.78
Dimethyl sulfide – Imidazole 2	140.89	93.21	103.84	83.55	88.29	71.48
Imidazole 1 – Imidazole 2	96.44	89.16	112.17	99.86	91.48	110.48
Acetamide – Ethyl thiolate	-	-	-	107.30	94.63	115.39
Acetamide – Dimethyl sulfide	-	-	-	172.36	174.55	174.34
Acetamide – Imidazole 1	-	-	-	88.46	85.30	99.33
Acetamide – Imidazole 2	-	-	-	88.94	86.86	103.33

Table 3.7 Binding energies (kcal/mol) of T1Cu1 and T1Cu2 ligands computed by MP2,

AMOEBA and AMOEBA-AOM.

	T1Cu1			**T1Cu2**		
	MP2	AMOEBA-AOM	AMOEBA	MP2	AMOEBA-AOM	AMOEBA
Ethyl thiolate	-230.0	-265.0	-231.1	-230.8	-219.9	-222.9
Dimethyl sulfide	-23.8	-22.0	-31.5	-8.5	-36.0	-14.7
Imidazole 1	-43.7	-54.8	-56.9	-43.0	-53.2	-56.9
Imidazole 2	-40.5	-48.9	-54.6	-43.0	-47.7	-31.4
Acetamide	-	-	-	-20.3	-45.3	-62.1

Table 3.8 Geometries of Cu^{2+} binding sites of 1AG6 and 1DYZ proteins obtained from

X-ray crystal structures and AMOEBA-AOM MD simulations.

	1AG6			1DYZ	
	Experimental	AMOEBA-AOM		Experimental	AMOEBA-AOM
Metal-ligand bond length (Å)					
CYS84	2.15	2.15 ± 0.04	CYS112	2.14	2.47 ± 0.09
MET92	2.88	2.85 ± 0.05	MET121	3.26	2.83 ± 0.05
HIS37	1.96	2.17 ± 0.05	HIS46	2.04	2.13 ± 0.05
HIS87	2.01	2.15 ± 0.05	HIS117	1.99	2.15± 0.05
			GLY45	2.72	2.50 ± 0.02
Ligand-metal-ligand angle (º)					
CYS84 – MET92	105.93	95.36 ± 4.30	CYS112 – MET121	105.27	103.42 ± 5.51
CYS84 – HIS37	129.91	122.98 ± 5.66	CYS112 – HIS46	132.56	136.32 ± 5.72
CYS84 – HIS87	120.07	133.83 ± 5.82	CYS112 – HIS117	121.05	115.66 ± 5.56
MET92 – HIS37	87.10	93.47 ± 5.20	MET121 – HIS46	73.89	79.92 ± 3.82
MET92 – HIS87	102.15	106.11 ± 6.00	MET121 – HIS117	88.34	92.61 ± 5.06
HIS37 – HIS87	103.04	95.81 ± 4.59	HIS46 – HIS117	106.39	107.10 ± 5.69
			GLY45 – CYS112	104.10	90.17 ± 5.21
			GLY45 – MET121	148.38	164.68 ± 4.15
			GLY45 – HIS46	77.77	86.38 ± 4.34
			GLY45 – HIS117	86.43	85.18 ± 4.62

Chapter 4. Summary

In the previous chapters, we have explained the importance of d-orbital electronic effects in describing the coordination chemistry of TM complexes. Two novel polarizable MM models derived from valence bond (VB) theory and the angular overlap model (AOM) were integrated into the AMOEBA force field, while maintaining a consistent treatment of classical electrostatic interactions. It is evident that both approaches can significantly improve the accuracy of AMOEBA for TM systems by explicitly accounting for the hypervalency and LF contributions. In this concluding section, we compare the AMOEBA-VB and AMOEBA-AOM models and discuss the future directions for improving their performances.

4.1 Comparison of the AMOEBA-VB and AMOEBA-AOM force fields

Although, VB theory and the AOM are based on different QM frameworks, they are ultimately complementary models seeking to describe the same QM phenomena. The key ideas of the VB theory are resonance and the donor-acceptor hybridization. Meanwhile, the AOM is derived from LF principle and is therefore closely related to traditional MO theory. The AMOEBA-VB model is chemically intuitive, and elements comprising its potential terms are straightforward. However, in its current iteration, AMOEBA-VB is not able to sufficiently handle the Jahn-Teller distortion without an explicit correction term. This should not be considered an inherent deficiency of the VB theory, but is most likely due to the incomplete nature of the resonance weighting function. On the other hand, although the implementation of the AMOEBA-AOM model is more elaborate, it can automatically handle the Jahn-Teller distortion and has demonstrated good accuracy and transferability for more complex TM systems.

Nevertheless, we feel both approaches are suitable basis for a general TM force field and can serve as starting points for further development.

4.2 Future Directions for AMOEBA-VB

As mentioned previously, an evident area of improvement for AMOEBA-VB is its treatment of Jahn-Teller distortion. The elongation of axial water molecules in octahedral Cu^{2+} complexes is not fully reproduced by the present AMOEBA-VB model. An explicit description based on a harmonic first order approximation improves the results. However, this necessarily requires the energy function to be applied selectively based on geometry and ligand type. As a result, this solution cannot provide a smooth transition for dynamic Jahn-Teller effect and exchange between hexa and penta-aqua coordination. A possible remedy is to explore alternative forms of the resonance weighting function to better describe the effects of Jahn-Teller distortion. This approach conforms to the tenets of VB theory and does not limit the generality of the model. Another aspect that can be explored is inclusion of explicit coupling between the molecular geometry and hybridization.

In addition to further optimization of the AMOEBA-VB aqueous TM model, future work should extend to more complex systems, especially cases with more than one type of ligand, as well as ligands with a significant π contribution. The VB model can be generalized for such systems through development of hybridization-mixing rules. Modifications to the resonance weighting function may also be necessary for these more complex situations. It is perhaps possible to adapt the π interaction terms from VALBOND[70] for the AMOEBA-VB model.

4.3 Future Directions for AMOEBA-AOM

There are two main areas of improvements that can be made to the AMOEBA-AOM formulation. First is a better method of handling the elongation of the dimethyl sulfide / methionine ligand as described earlier. A possible solution is by applying functional forms for e_{AOM} different from this initial iteration. Alternatively, a coupling of metal-ligand bonding to the L-M-L angle similar to the strategy of AMOEBA-VB[119] can be explored. A second aspect of the AMOEBA-AOM that can be improved is its accuracy in describing sulfur ligand binding energies. The Morse bonding term can be replaced with a different, more flexible, functional form. An interesting candidate is to reintroduce the buffered 14-7 vdW potential used by sthe standard AMOEBA force field for sulfur ligands since it shows remarkable agreement with QM energies. It should be noted that previous efforts to model the LF effects have been largely focused on geometries. We believe that the accurate description of ligand binding energies is also an important aspect of any MM model, especially if one wants to study ligand exchanges, vibrational frequencies and other dynamic events.

In addition to making improvements to the AMOEBA-AOM as outlined above and continuing refinement of the AMOEBA-AOM parameters, it would be interesting to apply the AMOEBA-AOM to other copper centers and produce a complete set of parameters for all amino acid ligands. One intriguing area for study is to investigateion of conformational changes in T1Cu proteins between their oxidized and reduced forms, Indeed, the two forms have different binding partners at the metal center. Since Cu^+ has a d^{10} configuration, it can be treated in similar fashion to Zn^{2+} as we have demonstrated[119], albeit with a different formal charge assignment. This work is planned for the near future.

Appendix A Derivatives of the AMOEBA-VB Potentials

The derivations are only presented with respect to X-coordinates since the forms for Y and Z-coordinates are very similar.
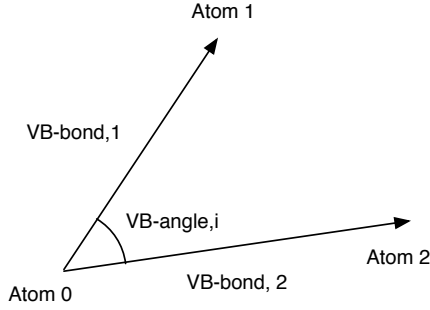
Pictorial representation of a single VB term

VB derivatives

$$E_{\text{vb}} = \sum_{i}^{\text{angles}} W_i(E_{\text{vb-bond,1}} + E_{\text{vb-bond,2}} + E_{\text{vb-angle},i})$$

$$\frac{\partial E_{\text{vb}}}{\partial x_0} = \sum_{i}^{\text{angles}} \frac{\partial W_i}{\partial x_0}(E_{\text{vb-bond,1}} + E_{\text{vb-bond,2}} + E_{\text{vb-angle},i}) + \sum_{i}^{\text{angles}} W_i\left(\frac{\partial E_{\text{vb-bond,1}}}{\partial x_0} + \frac{\partial E_{\text{vb-bond,2}}}{\partial x_0} + \frac{\partial E_{\text{vb-angle},i}}{\partial x_0}\right)$$

$$\frac{\partial E_{\text{vb}}}{\partial x_1} = \sum_{i}^{\text{angles}} \frac{\partial W_i}{\partial x_1}(E_{\text{vb-bond,1}} + E_{\text{vb-bond,2}} + E_{\text{vb-angle},i}) + \sum_{i}^{\text{angles}} W_i\left(\frac{\partial E_{\text{vb-bond,1}}}{\partial x_1} + \frac{\partial E_{\text{vb-angle},i}}{\partial x_1}\right)$$

$$\frac{\partial E_{\text{vb}}}{\partial x_2} = \sum_{i}^{\text{angles}} \frac{\partial W_i}{\partial x_2}(E_{\text{vb-bond,1}} + E_{\text{vb-bond,2}} + E_{\text{vb-angle},i}) + \sum_{i}^{\text{angles}} W_i\left(\frac{\partial E_{\text{vb-bond,2}}}{\partial x_2} + \frac{\partial E_{\text{vb-angle},i}}{\partial x_2}\right)$$

VB bond derivatives for a single VB angle

Chain rule for distance *r*

$$r_1 = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2}$$

$$\frac{\partial r_1}{\partial x_0} = \frac{-(x_1 - x0)}{\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2}} = \frac{-(x_1 - x_0)}{r_1}$$

$$\frac{\partial r_1}{\partial x_1} = \frac{x_1 - x0}{\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2}} = \frac{x_1 - x_0}{r_1}$$

$$r_2 = \sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2 + (z_2 - z_0)^2}$$

$$\frac{\partial r_2}{\partial x_0} = \frac{-(x_2 - x0)}{\sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2 + (z_2 - z_0)^2}} = \frac{-(x_2 - x_0)}{r_2}$$

$$\frac{\partial r_2}{\partial x_2} = \frac{x_2 - x0}{\sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2 + (z_2 - z_0)^2}} = \frac{x_2 - x_0}{r_2}$$

Derviatives for bond term

$$E_{\text{vb-bond},1} = k_{b1} e^{-a_{b1} r_1^2}$$

$$\frac{\partial E_{\text{vb-bond},1}}{\partial x_0} = -2k_{b1} a_{b1} r_1 e^{-a_{b1} r_1^2} \frac{\partial r_1}{\partial x_0}$$

$$\frac{\partial E_{\text{vb-bond},1}}{\partial x_1} = -2k_{b1} a_{b1} r_1 e^{-a_{b1} r_1^2} \frac{\partial r_1}{\partial x_1}$$

$$E_{\text{vb-bond},2} = k_2 e^{-a_2 r_2^2}$$

$$\frac{\partial E_{\text{vb-bond},1}}{\partial x_0} = -2k_{b2} a_{b2} r_2 e^{-a_{b2} r_2^2} \frac{\partial r_2}{\partial x_0}$$

$$\frac{\partial E_{\text{vb-bond},1}}{\partial x_2} = -2k_{b2} a_{b2} r_2 e^{-a_{b2} r_2^2} \frac{\partial r_2}{\partial x_2}$$

VB angle derivatives for a single VB angle

Chain rules for θ

$$\theta = \cos^{-1}\left(\frac{\vec{v_1} \cdot \vec{v_2}}{r_1 r_2}\right)$$

$$\vec{v_1} \cdot \vec{v_2} = (x_1 - x_0)(x_2 - x_0) + (y_1 - y_0)(y_2 - y_0) + (z_1 - z_0)(z_2 - z_0)$$

$$\frac{\partial \theta}{\partial x_0} = -\frac{1}{\sqrt{1 - (\frac{\vec{v_1} \cdot \vec{v_2}}{r_1 r_2})^2}} \cdot \frac{\partial(\frac{\vec{v_1} \cdot \vec{v_2}}{r_1 r_2})}{\partial x_0}$$

$$= -\frac{1}{\sqrt{1 - (\frac{\vec{v_1} \cdot \vec{v_2}}{r_1 r_2})^2}} \cdot \frac{\frac{\partial \vec{v_1} \cdot \vec{v_2}}{\partial x_0} r_1 r_2 - \vec{v_1} \cdot \vec{v_2} \frac{\partial r_1 r_2}{\partial x_0}}{(r_1 r_2)^2}$$

$$= -\frac{1}{\sqrt{1 - (\frac{\vec{v_1} \cdot \vec{v_2}}{r_1 r_2})^2}} \cdot \frac{[-(x_2 - x_0) - (x_1 - x_0)]r_1 r_2 - \vec{v_1} \cdot \vec{v_2}(\frac{\partial r_1}{\partial x_0} r_2 + \frac{\partial r_2}{\partial x_0} r_1)}{(r_1 r_2)^2}$$

$$= -\frac{1}{\sqrt{1 - (\frac{\vec{v_1} \cdot \vec{v_2}}{r_1 r_2})^2}} \cdot \frac{(2x_0 - x_1 - x_2)r_1 r_2 - \vec{v_1} \cdot \vec{v_2}(\frac{\partial r_1}{\partial x_0} r_2 + \frac{\partial r_2}{\partial x_0} r_1)}{(r_1 r_2)^2}$$

$$\frac{\partial \theta}{\partial x_1} = -\frac{1}{\sqrt{1 - (\frac{\vec{v_1} \cdot \vec{v_2}}{r_1 r_2})^2}} \cdot \frac{(x_2 - x_0)r_1 r_2 - \vec{v_1} \cdot \vec{v_2}(\frac{\partial r_1}{\partial x_1} r_2)}{(r_1 r_2)^2}$$

$$\frac{\partial \theta}{\partial x_2} = -\frac{1}{\sqrt{1 - (\frac{\vec{v_1} \cdot \vec{v_2}}{r_1 r_2})^2}} \cdot \frac{(x_1 - x_0)r_1 r_2 - \vec{v_1} \cdot \vec{v_2}(\frac{\partial r_2}{\partial x_2} r_1)}{(r_1 r_2)^2}$$

Derivatives for angular term

$$E_{\text{vb-angle,i}} = (k_{a1}e^{-a_{a1}r_1^2} + k_{a2}e^{-a_{a2}r_2^2})(1 - \Delta(\theta + \pi)^2)$$

$$\Delta = s + p\cos\theta + d\frac{3\cos^2\theta - 1}{2}$$

$$\frac{\partial \Delta(\theta + \pi)}{\partial x_0} = (p\sin\theta - 3d\cos\theta\sin\theta)\frac{\partial \theta}{\partial x_0}$$

$$\frac{\partial \Delta(\theta + \pi)}{\partial x_1} = (p\sin\theta - 3d\cos\theta\sin\theta)\frac{\partial \theta}{\partial x_1}$$

$$\frac{\partial \Delta(\theta + \pi)}{\partial x_2} = (p\sin\theta - 3d\cos\theta\sin\theta)\frac{\partial \theta}{\partial x_2}$$

$$\frac{\partial E_{\text{vb-angle,i}}}{\partial x_0} = (-2k_{a1}a_{a1}r_1 e^{-a_{a1}r_1^2}\frac{\partial r_1}{\partial x_0} - 2k_{a2}a_{a2}r_2 e^{-a_{a2}r_2^2}\frac{\partial r_2}{\partial x_0})(1 - \Delta(\theta + \pi)^2)$$
$$+ (k_{a1}e^{-a_{a1}r_1^2} + k_{a2}e^{-a_{a2}r_2^2})(-2\Delta(\theta + \pi)\frac{\partial \Delta(\theta + \pi)}{\partial x_0})$$

$$\frac{\partial E_{\text{vb-angle,i}}}{\partial x_1} = (-2k_{a1}a_{a1}r_1 e^{-a_{a1}r_1^2}\frac{\partial r_1}{\partial x_1})(1 - \Delta(\theta + \pi)^2)$$
$$+ (k_{a1}e^{-a_{a1}r_1^2} + k_{a2}e^{-a_{a2}r_2^2})(-2\Delta(\theta + \pi)\frac{\partial \Delta(\theta + \pi)}{\partial x_1})$$

$$\frac{\partial E_{\text{vb-angle,i}}}{\partial x_2} = (-2k_{a2}a_{a2}r_2 e^{-a_{a2}r_2^2}\frac{\partial r_2}{\partial x_2})(1 - \Delta(\theta + \pi)^2)$$
$$+ (k_{a1}e^{-a_{a1}r_1^2} + k_{a2}e^{-a_{a2}r_2^2})(-2\Delta(\theta + \pi)\frac{\partial \Delta(\theta + \pi)}{\partial x_2})$$

VB derivatives for resonance weighting function W

$$W_i = C_i e^{-a_{i1}r_{i1}^2}e^{-a_{i2}r_{i2}^2}(b_i + \sum_j^{\text{angles}} c_j e^{-a_{j1}r_{j1}^2}e^{-a_{j2}r_{j2}^2})$$

$$\frac{\partial W_i}{\partial x_0} = -2(a_{i1}r_{i1}\frac{\partial r_{i1}}{\partial x_0} + a_{i2}r_{i2}\frac{\partial r_{i2}}{\partial x_0})W_i$$
$$+ C_i e^{-a_{i1}r_{i1}^2}e^{-a_{i2}r_{i2}^2}\sum_j^{\text{angles}} -2c_j(a_{j1}r_{j1}\frac{\partial r_{j1}}{\partial x_0} + a_{j2}r_{j2}\frac{\partial r_2}{\partial x_0})e^{-a_{j1}r_{j1}^2}e^{-a_{j2}r_{j2}^2}$$

$$\frac{\partial W_i}{\partial x_1} = -2(a_{i1}r_{i1}\frac{\partial r_{i1}}{\partial x_1})W_i + C_i e^{-a_{i1}r_{i1}^2}e^{-a_{i2}r_{i2}^2}\sum_{j=\text{atom 1 of }i}^{\text{angles}} -2c_j(a_{j1}r_{j1}\frac{\partial r_{j1}}{\partial x_1})e^{-a_{j1}r_{j1}^2}e^{-a_{j2}r_{j2}^2}$$

$$\frac{\partial W_i}{\partial x_2} = -2(a_{i2}r_{i2}\frac{\partial r_{i2}}{\partial x_2})W_i + C_i e^{-a_{i1}r_{i1}^2}e^{-a_{i2}r_{i2}^2}\sum_{j=\text{atom 2 of }i}^{\text{angles}} -2c_j(a_{j2}r_{j2}\frac{\partial r_{j2}}{\partial x_2})e^{-a_{j1}r_{j1}^2}e^{-a_{j2}r_{j2}^2}$$

Appendix B AMOEBA Parameters for TM Ligands

Parameters are reported in standard TINKER key file format.

```
##################
# Ethyl thiolate #
##################


atom          411  411   C   "Thiolate CH3  "      6    12.011   4

atom          412  412   H   "Thiolate H3C  "      1     1.008   1

atom          413  413   C   "Thiolate CH2S-"      6    12.011   4

atom          414  414   H   "Thiolate H2CS-"      1     1.008   1

atom          415  415   S   "Thiolate S-   "     16    32.066   1


polarize      411         1.3340    0.3900    412    413

polarize      412         0.4960    0.3900    411

polarize      413         1.3340    0.3900    411    414     415

polarize      414         0.4960    0.3900    413

polarize      415         4.0000    0.3900    413


multipole   411  413  412            -0.20281

                                     -0.00301   0.00000    0.21905

                                     -0.12856

                                      0.00000  -0.17155

                                     -0.18498   0.00000    0.30011

multipole   412  411  413             0.02125

                                      0.03105   0.00000   -0.07061

                                      0.08689
```

				0.00000	0.06296	
				0.04663	0.00000	-0.14985
multipole	413	415	411	0.25613		
				0.27834	0.00000	0.39573
				0.14308		
				0.00000	-0.26550	
				-0.01686	0.00000	0.12242
multipole	414	413	415	-0.03752		
				0.07455	0.00000	-0.05598
				0.04192		
				0.00000	0.00212	
				0.05471	0.00000	-0.04404
multipole	415	413	411	-1.04203		
				0.02076	0.00000	0.33904
				-1.57307		
				0.00000	-1.72462	
				0.01309	0.00000	3.29769

vdw	411		3.8200	0.1040	
vdw	412		2.9800	0.0240	0.920
vdw	413		3.7800	0.1060	
vdw	414		2.8700	0.0330	0.900
vdw	415		4.3500	0.3850	

bond	411	412	341.00	1.1120
bond	411	413	323.00	1.5247
bond	413	414	395.50	1.0850
bond	413	415	216.00	1.8225

angle	411	413	415		53.20	112.73		
angle	414	413	415		35.00	109.98		
angle	412	411	412		39.57	107.60	107.80	109.47
angle	412	411	413		42.44	109.80	109.31	110.70
angle	414	413	411		42.44	109.80	109.31	110.70
angle	414	413	414		39.57	107.60	107.80	109.47

strbnd	411	413	415	18.70	18.70
strbnd	414	413	415	11.50	11.50
strbnd	414	413	411	11.50	11.50
strbnd	413	415	413	-5.75	-5.75
strbnd	413	415	415	-5.75	-5.75
strbnd	413	411	412	11.50	11.50

torsion   415 413 411 412   0.000 0.0 1   0.000 180.0 2   0.475
0.0 3

torsion   414 413 411 412   0.000 0.0 1   0.000 180.0 2   0.238
0.0 3

############

Phenoxide

############

atom	421	421	C	"Phenoxide C-ortho"	6	12.011	3
atom	422	421	C	"Phenoxide C-meta"	6	12.011	3
atom	423	421	C	"Phenoxide C-para"	6	12.011	3
atom	424	421	C	"Phenoxide C-O-"	6	12.011	3
atom	425	422	H	"Phenoxide H-ortho"	1	1.008	1
atom	426	422	H	"Phenoxide H-meta"	1	1.008	1

```
atom      427  422    H     "Phenoxide H-para"        1      1.008      1

atom      428  423    O     "Phenoxide O-"            8     15.999      1


vdw       421                3.8000      0.0910

vdw       422                2.9800      0.0260      0.920

vdw       423                3.3200      0.1120


bond      421  421          210.90      1.3650

bond      421  422          409.50      1.0800

bond      421  423          680.00      1.2747


angle     421  421  421      64.67     121.70

angle     421  421  422      35.25     120.00     120.50        0.00

angle     421  421  423      60.00     123.57


strbnd    421  421  421      18.70      18.70

strbnd    421  421  422      38.00      11.60

strbnd    421  421  423      18.70      18.70


opbend    421  421   0    0              14.40

opbend    422  421   0    0              15.10

opbend    423  421   0    0              14.40


torsion   421  421  421  421     -0.670 0.0 1   4.004 180.0 2    0.000
0.0 3

torsion   421  421  421  422      0.550 0.0 1   4.534 180.0 2   -0.550
0.0 3

torsion   422  421  421  422      0.000 0.0 1   4.072 180.0 2    0.000
0.0 3
```

116

```
torsion     421 421 421  423      0.000 0.0 1   4.470 180.0 2    0.000
0.0 3

torsion     422 421 421  423      0.000 0.0 1   4.470 180.0 2    0.000
0.0 3


polarize    421          1.7500   0.3900      422    424    425

polarize    422          1.7500   0.3900      421    423    426

polarize    423          1.7500   0.3900      422    427

polarize    424          1.7500   0.3900      421    428

polarize    425          0.6960   0.3900      421

polarize    426          0.6960   0.3900      422

polarize    427          0.6960   0.3900      423

polarize    428          0.8370   0.3900      424


multipole   421 -424 -422            -0.23130

                                      0.41299    0.00000   -0.00719

                                     -0.84532

                                      0.00000   -0.28695

                                     -0.22512    0.00000    1.13227
multipole   422 -421 -423             0.09326

                                      0.10628    0.00000    0.23688

                                     -0.36832

                                      0.00000   -0.69071

                                      0.27592    0.00000    1.05903
multipole   423 -422 -422            -0.15839

                                      0.00000    0.00000   -0.16859

                                      0.58328

                                      0.00000   -0.64580

                                      0.00000    0.00000    0.06252
```

```
multipole   424  428  421              0.63120
                                       0.00000    0.00000    0.27846
                                       0.62275
                                       0.00000   -0.12881
                                       0.00000    0.00000   -0.49394
multipole   425  421  422             -0.04632
                                       0.26641    0.00000   -0.11778
                                      -0.37553
                                       0.00000    0.16803
                                       0.37648    0.00000    0.20750
multipole   426  422  423             -0.05729
                                       0.10056    0.00000   -0.29672
                                       0.36124
                                       0.00000   -0.01213
                                       0.22601    0.00000   -0.34911
multipole   427  423  422             -0.05751
                                       0.00000    0.00000   -0.25877
                                       0.24169
                                       0.00000   -0.04974
                                       0.00000    0.00000   -0.19195
multipole   428  424  421             -0.93199
                                       0.00000    0.00000    0.04254
                                       0.22122
                                       0.00000   -0.08151
                                       0.00000    0.00000   -0.13971
```

Appendix C Additional Data for T1Cu1 and T1Cu2

Comparisons of QM, AMOEBA and AMOEBA-AOM potentials for randomly perturbed T1Cu1 structures; results obtained by AMOEBA are on the left and those by AMOEBA-AOM are on the right; only a single ligand is perturbed from the QM optimized geometry in each plot.

Correlation between AMOEBA and MP2 potentials (T1Cu1/imidazole 1)

y = 0.24685x + 1.337

R-squared = 0.090178

RMS error = 1.7871

data = 90

Correlation between AMOEBA-AOM and MP2 potentials (T1Cu1/imidazole 1)

y = 1.0797x + 0.07985

R-squared = 0.91619

RMS error = 0.44648

data = 90

Correlation between AMOEBA and MP2 potentials (T1Cu1/imidazole 2)

y = 0.20079x + 1.5978

R-squared = 0.10244

RMS error = 2.789

data = 83

Correlation between AMOEBA-AOM and MP2 potentials (T1Cu1/imidazole 2)

y = 1.0722x + 0.22745

R-squared = 0.81357

RMS error = 0.56244

data = 83
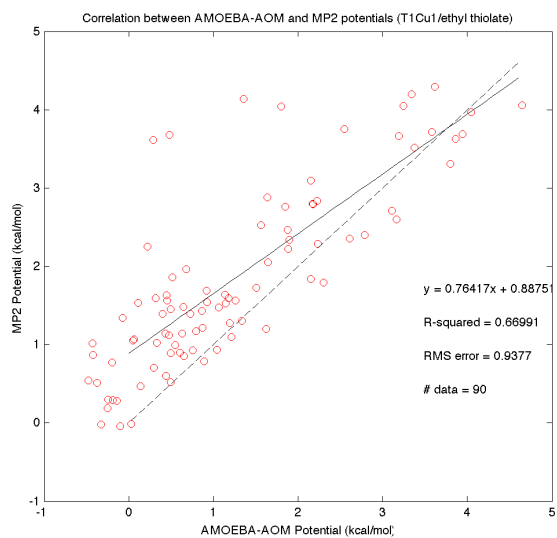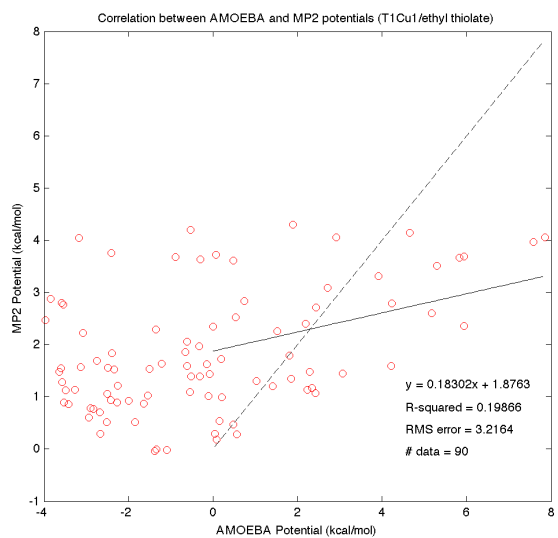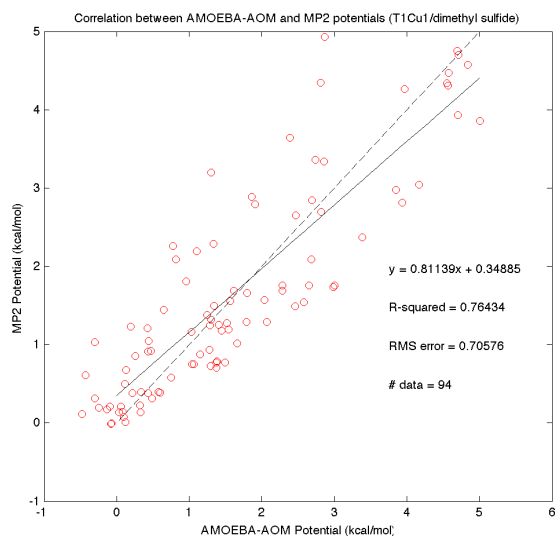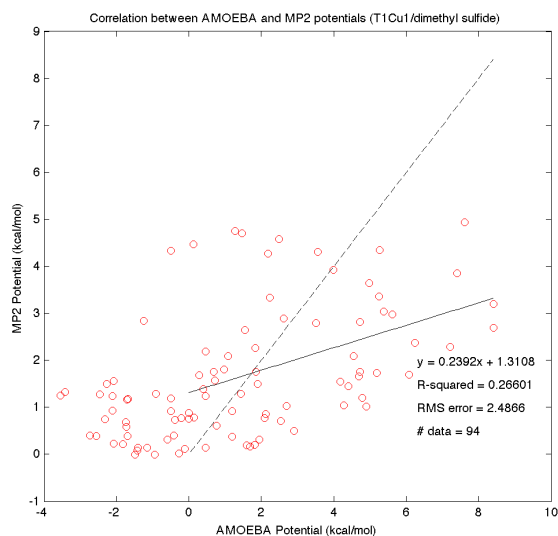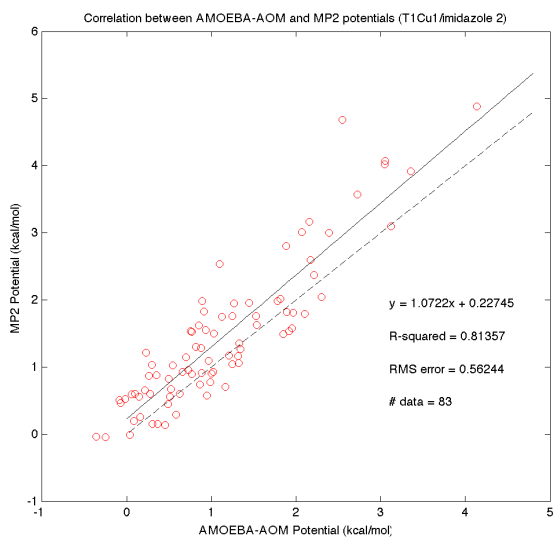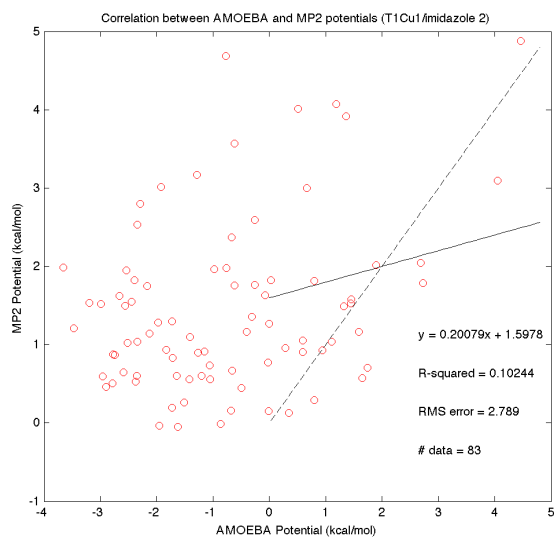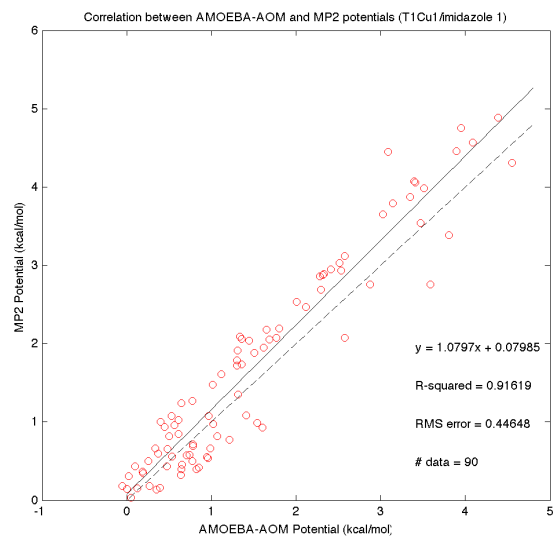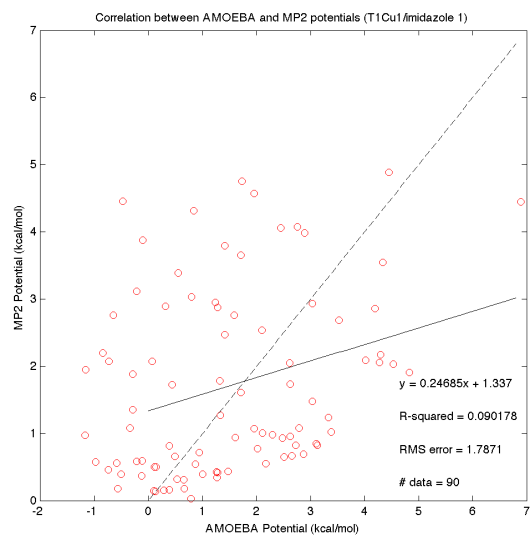
Comparisons of QM, AMOEBA and AMOEBA-AOM potentials for randomly perturbed T1Cu2 structures; results obtained by AMOEBA are on the left and those by AMOEBA-AOM are on the right; only a single ligand is perturbed from the QM optimized geometry in each plot.

Correlation between AMOEBA and MP2 potentials (T1Cu2/imidazole 1)

y = 0.34245x + 1.1114

R-squared = 0.27718

RMS error = 1.8319

data = 93

Correlation between AMOEBA-AOM and MP2 potentials (T1Cu2/imidazole 1)

y = 1.171x + 0.20135

R-squared = 0.88848

RMS error = 0.6463

data = 93

Correlation between AMOEBA and MP2 potentials (T1Cu2/imidazole 2)

y = 0.76723x + 1.1397

R-squared = 0.56972

RMS error = 1.327

data = 79

Correlation between AMOEBA-AOM and MP2 potentials (T1Cu2/imidazole 2)

y = 0.78443x + 0.8002

R-squared = 0.8269

RMS error = 0.83929

data = 79

Correlation between AMOEBA and MP2 potentials (T1Cu2/acetamide)

y = 0.89735x + 0.86094

R-squared = 0.93228

RMS error = 0.79436

data = 76

Correlation between AMOEBA-AOM and MP2 potentials (T1Cu2/acetamide)

y = 0.72108x + 0.28964

R-squared = 0.98958

RMS error = 0.93074

data = 76

References

(1)   Lippard, S. J.; Berg, J. M. *Principles of Bioinorganic Chemistry*; University Science Books, 1994.

(2)   Holm, R. H.; Kennepohl, P.; Solomon, E. I. *Chem. Rev.* **1996**, *96*, 2239–2314.

(3)   *Bioorganometallics: Biomolecules, Labeling, Medicine*; Jaouen, G., Ed. Wiley-VCH: Weinheim, 2006.

(4)   *Biological Inorganic Chemistry: Structure and Reactivity*; Gray, H. B.; Stiefel, E. I.; Valentine, J. S.; Bertini, I., Eds. 1st ed. University Science Books, 2006.

(5)   Crabtree, R. H. *The Organometallic Chemistry of the Transition Metals*; 5 ed. John Wiley & Sons, Inc., 2009.

(6)   Hartinger, C. G.; Dyson, P. J. *Chem. Soc. Rev.* **2009**, *38*, 391.

(7)   Hillard, E. A.; Jaouen, G. *Organometallics* **2011**, *30*, 20–27.

(8)   Klug, A. *Annu. Rev. Biochem.* **2010**, *79*, 213–231.

(9)   Rulíšek, L.; Jensen, K. P.; Lundgren, K.; Ryde, U. *J. Comput. Chem.* **2006**, *27*, 1398–1414.

(10)   Srnec, M.; Aquilante, F.; Ryde, U.; Rulíšek, L. *J. Phys. Chem. B* **2009**, *113*, 6074–6086.

(11)   Bertini, I.; Gray, H. B.; Stiefel, E. I.; Valentine, J. S. *Biological Inorganic Chemistry: Structure & Reactivity*; University Science Books: Sausalito, 2007.

(12)   Constable, E. G.; Gerloch, M. *Transition Metal Chemistry*; VCH Verlagsgesellschaft mbH, Weinheim, 1997.

(13)   *Concepts and Models in Bioinorganic Chemistry*; Kraatz, H.-B.; Metzler-Nolte, N., Eds. Wiley-VCH: Weinheim, 2006.

(14)   Jean, Y. *Molecular Orbitals of Transition Metal Complexes*; Oxford University Press, USA, 2005.

(15)   Jahn, H.; Teller, E. *Proceedings of the Royal Society of London Series A* **1937**, *161*, 220–235.

(16)   Keinan, S.; Avnir, D. *Inorg. Chem.* **2001**, *40*, 318–323.

(17)   Jensen, F. *Introduction to Computational Chemistry*; 2nd ed. John Wiley & Sons, Inc.: West Sussex, 2007.

(18)   Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389–427.

(19)   Harvey, J. N. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **2006**, *102*, 203.

(20)   Cramer, C. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10757.

(21)   Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(22)   Becke, A. D. *Physical Review A* **1988**, *38*, 3098.

(23)   Lee, C.; Yang, W.; Parr, R. *Phys. Rev., B Condens. Matter* **1988**, *37*, 785–789.

(24)   Grimme, S. *J. Chem. Phys.* **2006**.

(25)   Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**.

(26)   Yang, W.; Lee, T.-S. *J. Chem. Phys.* **1995**, *103*, 5674–5678.

(27)   Ochsenfeld, C.; White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **1998**, *109*, 1663–1669.

(28)   Babu, K.; Gadre, S. R. *J. Comput. Chem.* **2003**, 484–495.

(29)   Winget, P.; Sel uki, C.; Horn, A. H. C.; Martin, B.; Clark, T. *Theor Chem Acc* **2003**, *110*, 254–266.

(30)   Bredow, T.; Jug, K. *Theor Chem Acc* **2005**, *113*, 1–14.

(31)   Wu, X.; Koslowski, A.; Thiel, W. *J. Chem. Theory Comput.* **2012**, *8*, 2272–2281.

(32)   Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *21*, 1442–1457.

(33)   Sanz-Navarro, C. F.; Grima, R.; García, A.; Bea, E. A.; Soba, A.; Cela, J. M.; Ordejón, P. *Theor Chem Acc* **2010**, *128*, 825–833.

(34)   Lin, H.; Truhlar, D. G. *Theor Chem Acc* **2006**, *117*, 185–199.

(35)   Weinhold, F.; Landis, C. R. *Valency and bonding*; Cambridge Univ Pr, 2005.

(36)   Weinhold, F. *J. Comput. Chem.* **2012**, *33*, 2363–2379.

(37)   Antonyuk, S. V.; Strange, R. W.; Marklund, S. L.; Hasnain, S. S. *J. Mol. Biol.* **2009**, *388*, 310–326.

(38)   Paoli, M.; Liddington, R.; Tame, J.; Wilkinson, A.; Dodson, G. *J. Mol. Biol.* **1996**, *256*, 775–792.

(39)   Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 1987.

(40)   Leach, A. *Molecular Modelling: Principles and Applications*; 2nd ed. Prentice Hall: Edinburgh Gate, 2001.

(41)   Comba, P.; Hambley, T. W. *Molecular Modeling of Inorganic Compounds*; 2nd ed. Wiley-VCH, 2001.

(42)   Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; Academic Press: San Diego, 2002.

(43)   Comba, P.; Hambley, T. W.; Martin, B. *Molecular Modeling of Inorganic Compounds*; Wiley-VCH, 2009.

(44)   Morse, P. M. *Physical Review* **1929**, *34*, 57.

(45)   Jones, J. E. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **1924**, *106*, 463–477.

(46)   Halgren, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 7827–7843.

(47)   Buckingham, R. A. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **1938**, *168*, 264–283.

(48)   Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.;

Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(49)   Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(50)   Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(51)   Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(52)   Stone, A. *J. Chem. Theory Comput.* **2005**.

(53)   Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

(54)   Piquemal, J.-P.; Perera, L.; Cisneros, G. A.; Ren, P.; Pedersen, L. G.; Darden, T. A. *J. Chem. Phys.* **2006**, *125*, 054511.

(55)   Hodges, M. P.; Stone, A. J.; Xantheas, S. S. *J. Phys. Chem. A* **1997**, *101*, 9163–9168.

(56)   Patel, S.; Mackerell, A. D.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1504–1514.

(57)   Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; Mackerell, A. D. *J. Chem. Theory Comput.* **2005**, *1*, 153–168.

(58)   Ren, P.; Ponder, J. W. *J. Comput. Chem.* **2002**, *23*, 1497–1506.

(59)   *pentakis(λ3-oxidanylidynemethyl)iron*; ChemSpider.

(60)   Nielson, K. D.; van Duin, A. C. T.; Oxgaard, J.; Deng, W.-Q.; Goddard, W. A. *J. Phys. Chem. A* **2005**, *109*, 493–499.

(61)   van Duin, A. C. T.; Bryantsev, V. S.; Diallo, M. S.; Goddard, W. A.; Rahaman, O.; Doren, D. J.; Raymand, D.; Hermansson, K. *J. Phys. Chem. A* **2010**, *114*, 9507–9514.

(62)   Pauling, L. *Physical Review* **1938**, *54*, 899–904.

(63)   Pauling, L. *Proceedings of the Royal Society of London Series A* **1949**, *196*, 343–362.

(64)   Pauling, L. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 4200–4202.

(65)   Pauling, L. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 1403–1405.

(66)   Schäffer, C. E.; Jørgensen, C. K. *Molecular Physics* **1965**, *9*, 401–412.

(67)   Landis, C.; Cleveland, T. *J. Am. Chem. Soc.* **1993**, *115*, 4201–4209.

(68)   Cleveland, T.; Landis, C. R. *J. Am. Chem. Soc.* **1996**, *118*, 6020–6030.

(69)   Landis, C.; Cleveland, T. *J. Am. Chem. Soc.* **1998**, *120*, 2641–2649.

(70)   Firman, T. K.; Landis, C. R. *J. Am. Chem. Soc.* **2001**, *123*, 11728–11742.

(71)   Deeth, R. J.; Foulis, D. L. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4292–4297.

(72)   Deeth, R. J. *Faraday Disc.* **2003**, *124*, 379.

(73)   Deeth, R. J.; Fey, N.; Williams-Hubbard, B. *J. Comput. Chem.* **2004**, *26*, 123–130.

(74)   Piquemal, J.-P.; Williams-Hubbard, B.; Fey, N.; Deeth, R. J.; Gresh, N.; Giessner-Prettre, C. *J. Comput. Chem.* **2003**, *24*, 1963–1970.

(75)   Carlsson, A. E. *Phys. Rev. Lett.* **1998**, *81*, 477–480.

(76)   Carlsson, A. E.; Zapata, S. *Biophysical Journal* **2001**, *81*, 1–10.

(77)   Zapata, S.; Carlsson, A. *Phys. Rev. B* **2002**, *66*.

(78)   Carlsson, H.; Haukka, M.; Bousseksou, A.; Latour, J.-M.; Nordlander, E. *Inorg. Chem.* **2004**, *43*, 8252–8262.

(79)   Ponder, J. W.; Case, D. A. *Advances in protein chemistry* **2003**, *66*, 27–85.

(80)   Grossfield, A.; Ren, P.; Ponder, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 15671–15682.

(81)   Jiao, D.; King, C.; Grossfield, A.; Darden, T. A.; Ren, P. *J. Phys. Chem. B* **2006**, *110*, 18553–18559.

(82)   Binning, R. C.; Curtiss, L. A. *J. Comput. Chem.* **1990**, *11*, 1206–1216.

(83)   Porterfield, W. *Inorganic Chemistry: A Unified Approach*; Academic Pr, 1993.

(84)   Wu, J. C.; Piquemal, J.-P.; Chaudret, R.; Reinhardt, P.; Ren, P. *J. Chem. Theory Comput.* **2010**, *6*, 2059–2070.

(85)   Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(86)   Ren, P.; Wu, C.; Ponder, J. W. *J. Chem. Theory Comput.* **2011**, *7*, 3143–3161.

(87)   Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, 1997.

(88)   Thole, B. T. *Chemical Physics* **1981**, *59*, 341–350.

(89)   Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2004**, *208*, 13427–13437.

(90)   Deeth, R. J.; Hitchman, M. A. *Inorg. Chem.* **1986**, *25*, 1225–1233.

(91)   Bersuker, I. B. *Chem. Rev.* **2001**, *101*, 1067–1114.

(92)   Comba, P.; Zimmer, M. *Inorg. Chem.* **1994**, *33*, 5368–5369.

(93)   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, N.; Vreven, T.; Montgomery, J. A., Jr; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision A.2 **2009**.

(94)   Dunning, T. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(95)   Balabanov, N. B.; Peterson, K. A. *J. Chem. Phys.* **2005**, *123*, 064107.

(96)   Rabuck, A. D.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 695–700.

(97)   Ohtaki, H. *Monatshefte für Chemie* **2001**, *132*, 1237–1268.

(98)   Nymand, T. M. *J. Chem. Phys.* **2000**, *112*, 6152–6160.

(99)   Toukmaji, A.; Sagui, C.; Board, J.; Darden, T. *J. Chem. Phys.* **2000**, *113*, 10913.

(100)   Basolo, F.; Pearson, R. G. *Mechanisms of Inorganic Reactions: A Study of Metal Complexes in Solution*; 2nd ed. Wiley New York, 1958.

(101)   Hartmann, M.; Clark, T. *J. Am. Chem. Soc.* **1997**, *119*, 7843–7850.

(102)   de Almeida, K. J.; Murugan, N. A.; Rinkevicius, Z.; Hugosson, H. W.; Vahtras, O.; Ågren, H.; Cesar, A. *Phys. Chem. Chem. Phys.* **2009**, *11*, 508.

(103)   Kumar, R.; Keyes, T. *J. Am. Chem. Soc.* **2011**, *133*, 9441–9450.

(104)   Helm, L.; Merbach, A. E. *Coordination Chemistry Reviews* **1999**, *187*, 151–181.

(105)   Pasquarello, A.; Petri, I.; Salmon, P. S.; Parisel, O.; Car, R.; Toth, E.; Powell, D. D.; Fischer, H. E.; Helm, L.; Merbach, A. E. *Science* **2001**, *291*, 856–859.

(106)   Rode, B. M.; Schwenk, C. F.; Hofer, T. S.; Randolf, B. R. *Coordination Chemistry Reviews* **2005**, *249*, 2993–3006.

(107)   Neilson, G. W.; Newsome, J. R.; Sandström, M. *J. Chem. Soc., Faraday Trans. 2* **1981**, *77*, 1245–1256.

(108)   Sham, T.; Hastings, J.; Perlman, M. *Chemical Physics Letters* **1981**, *83*, 391–396.

(109)   Beagley, B.; Eriksson, A.; Lindgren, J.; Persson, I.; Pettersson, L. G. M.; Sandstrom, M.; Wahlgren, U.; White, E. W. *J. Phys.: Condens. Matter* **1989**, *1*, 2395–2408.

(110)   Amira, S.; Spångberg, D.; Hermansson, K. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2874.

(111)   Xue, Y.; Okvist, M.; Hansson, O.; Young, S. *Protein Sci.* **1998**, *7*, 2099–2105.

(112)   Dodd, D. E.; Abraham, D. H. L.; Eady, D. R.; Hasnain, S. S. *Acta Cryst (2000). D56, 690-696 [doi:10.1107/S0907444900003309]* **2000**, 1–7.

(113)   De Rienzo, F.; Gabdoulline, R. R.; Menziani, M. C.; Wade, R. C. *Protein Sci.* **2000**, *9*, 1439–1454.

(114)   Comba, P.; Remenyi, R. *J. Comput. Chem.* **2002**, *23*, 697–705.

(115)   De Rienzo, F.; Gabdoulline, R. R.; Wade, R. C.; Sola, M.; Menziani, M. C. *Cell. Mol. Life Sci.* **2004**, *61*, 1123–1142.

(116)   Deeth, R. J. *Inorg. Chem.* **2007**, *46*, 4492–4503.

(117)   Figgis, B. N.; Hitchman, M. A. *Ligand Field Theory And Its Applications*; Wiley-VCH: New York, 2000.

(118)   Rose, M. *Elementary Theory Of Angular Momentum*; Wiley: New York, 1957.

(119)   Xiang, J. Y.; Ponder, J. W. *J. Comput. Chem.* **2012**, *34*, 739–749.

(120)   Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chemical Physics Letters* **1988**.

(121)   Uranga, J.; Mikulskis, P.; Genheden, S.; Ryde, U. *Computational and Theoretical Chemistry* **2012**, *1000*, 75–84.

(122)   Swift, T. J.; Connick, R. E. *J. Chem. Phys.* **1962**, *37*, 307–320.

(123)   Ohtaki, H.; Radnai, T. *Chem. Rev.* **1993**, *93*, 1157–1204.

(124)   Salmon, P. S.; Bellissent-Funel, M.-C.; Herdman, G. J. *Journal of Physics: Condensed Matter* **1990**, *2*, 4297.

(125)   Deeth, R. J.; Anastasi, A.; Diedrich, C.; Randell, K. *Coordination Chemistry Reviews* **2009**, *253*, 795–816.