Development of Semiempirical Models for Metalloproteins

Bharat Kumar Sharma

A Thesis

In

The Department

of

Physics

Presented in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy (Physics) at

Concordia University

Montreal, Quebec, Canada

August 2016

© Bharat Kumar Sharma, 2016

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Bharat Kumar Sharma

Entitled: Development of Semiempirical Models for Metalloproteins

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Physics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. D. Walsh

_____ External Examiner
Dr. N.J. Mosey

_____ External to Program
Dr. G. Peslherbe

_____ Examiner
Dr. L. Kalman

_____ Examiner
Dr. V. Zazubovits

_____ Thesis Supervisor
Dr. G. Lamoureux

Approved by: _____
Dr. P. Bianucci, Graduate Program Director

August 29, 2016 _____
Dr. A. Roy, Dean, Faculty of Arts and Science

Abstract

Development of Semiempirical Models for Metalloproteins

Bharat Kumar Sharma, Ph.D.

Concordia University, 2016

Theoretical models and computational techniques are useful for gaining insight into the interactions, movements, and functions of atoms and molecules, ranging from small chemical systems with few atoms to large biological molecules with many atoms. Due to the inability of force field methods to accurately describe different properties of metalloenzymes and the prohibitive computing cost of high-level quantum methods, computationally efficient models are needed.

This dissertation describes the development of new quantum semiempirical models for metalloproteins. The original AM1 (Austin Model 1) based on the neglect of diatomic differential overlap approximations was re-parameterized to describe the structural and energetic properties of biomolecules that mimic the active sites of metalloproteins. The biologically inspired genetic algorithm PIKAIA was used to optimize the parameters for each chemical element. Structures and energies of various clusters analogous to complexes found in metalloproteins were prepared as a training set using hybrid density functional theory. Models were trained to reproduce all of the properties included in the small training set. The optimized models were validated for large testing sets that incorporate bigger complexes and related reactions. Finally, the optimized models were used to study biologically-relevant processes in condensed phase using molecular dynamics simulations. All the gas- and liquid-phase results from the optimized models were compared with original semiempirical models as well as available high-level theoretical and experimental results.

Metal ions play crucial roles in biological systems. They actively participate in structural, catalytic, and co-catalytic activities of a large number of enzymes. The development of semiempirical models is divided into three parts. First, new AM1 parameters for hydrogen and oxygen were developed to describe gas-phase proton transfer reactions in water and static and dynamic properties of liquid water. Gas-phase results were compared with original AM1, RM1, and PM3 models, whereas liquid results were compared with original AM1, AM1-W, and

AM1PG-W models, and with available experimental results. It is found that the optimized model reproduces experimental data better than other available semiempirical models. Second, using the previously optimized model for hydrogen and oxygen, the AM1 model is re-parameterized for zinc and sulfur to describe important physical and chemical properties of zinc, water, hydrogen sulfide complexes mimicking structural motifs found in zinc enzymes. Metal-induced pK_a shifts are computed for water and hydrogen sulfide, and compared with available theoretical and experimental results. Third, using previously optimized parameters for hydrogen, oxygen, and zinc, AM1 parameters for carbon and nitrogen are optimized to study proton transfer, nucleophilic attacks, and peptide hydrolysis mechanisms in zinc metalloproteases.

Overall, the optimized models give promising results for the various properties of biomolecules in gas-phase clusters and in condensed phase. Particularly, the water model reproduces the proton transfer related properties in gas-phase and the structure, dielectric properties, and infrared spectra of liquid water. The zinc/sulfur model reproduces the hydration structure of zinc cation and zinc-bound hydrogen sulfide. Results for the coordination configurations of zinc solvated in water and in hydrogen sulfide confirm the versatility of the model. The optimized model for carbon and nitrogen improves the overall performance compared to AM1 and PM3. The optimized model for carbon and nitrogen reproduces structures and various energetic terms for zinc-ligands systems (representing the active sites of zinc enzymes) when compared to density functional theory results. The optimized model can be used to study metal-ligand reactivity in zinc enzymes.

Acknowledgements

All the work presented in this dissertation was done under the affiliations of Department of Physics and Centre for Research in Molecular Modeling (CERMM) at Concordia University, Montreal, Quebec, Canada during my Ph.D. studies from May 2012 to August 2016. This dissertation would not have been possible without the direct and indirect contributions, encouragements, and support of many people and organizations.

I would like to show my deepest gratitude to my supervisor Dr. Guillaume Lamoureux for the encouragements, supervision, and support since the beginning of my Ph.D. program. I am very grateful for having received the opportunity to work in his research lab and for having been allowed to work on my own schedule. Dr. Lamoureux helped me to understand how applications of theoretical models are very useful to understand microscopic behavior of macroscopic systems. His inspiration, new ideas in molecular modeling, and passions for science helped me to enjoy the course of this study. I am also thankful to Dr. Lamoureux for providing financial support in the forms of research bursary and conference travel grant during the entire duration of my studies.

I gratefully acknowledge my Ph.D. committee members, Dr. Laszlo Kalman and Dr. Valter Zazubovits, for valuable discussions, comments, and advice. I would also like to thank Marie-Anne Cheong Youne, Assistant to the Chair, Department of Physics, for providing valuable information and unflagging support to settle down in the city.

I would like to thank Dr. Shihao Wang for helping me to understand parameterization techniques and molecular dynamics simulations and for the valuable suggestions. I would also like to thank all my colleagues in the Department of Physics and CERMM. Special thanks go to Dr. Esam Orabi for his sincere help with many ab initio calculations and to colleagues Dr. Mariya Shadrina, Dr. Xijun Wang, Vinod Parmar, Tugba Nur Ozturk, Felix Pigeon, Jennifer Garner, Philippe Archambault, and Robert Ojike Wosu for their direct and indirect help during my studies. I would also like to thank Dr. Garold Murdachaew for providing the modified CP2K codes and for helpful discussion, and Dr. Leonardo Bernasconi for the pK_a calculation scripts and for valuable suggestions.

This would not have been possible without the financial support from many organizations. I would like to acknowledge all the funding sources for supporting me during my Ph.D. studies. Particularly I gratefully acknowledge Concordia University for “Concordia University Full-Tuition International Recruitment Award” and “Faculty of Arts and Science Graduate Fellowship” and PROTEO for “PROTEO Graduate Students Scholarship”. I would also like to acknowledge CERMM, Compute Canada, Calcul Quebec, and WestGrid for the computational resources.

Finally, I would like to thank my parents, grandparents, and other family members for their love, encouragement, and support in my entire life. My mother, Mrs. Durga Devi Sharma, is in the first place, for always giving me unconditional love and always being on my side since my childhood. Many thanks to my wife Mrs. Rakchhya Niraula for her love, faithful support, and encouragement. Thank you to my brother Mr. Suman Kumar Sharma for taking care of the parents. I would like to thank Mr. Donald Roy and Mrs. Lorna Roy for their generous love and assistance during the difficult times in Montreal. I am also grateful for all my friends inside and outside the Concordia family, who played important roles during my studies. I apologize to all for not mentioning their names one by one.

Bharat Kumar Sharma
Concordia University
August 2016

This dissertation is dedicated to my beloved mother Mrs. Durga Devi Sharma.

Table of Contents

List of Figures

List of Tables

Abbreviations

AM1	Austin Model 1
Asp	Aspartate
B3LYP	Becke, Three-parameter, Lee-Yang-Parr
CA	Carbonic anhydrase
CRF	Core repulsion function
CSD	Cambridge structural database
Cys	Cysteine
DFT	Density functional theory
DNA	Deoxyribonucleic acid
GA	Genetic algorithm
GGA	Generalized-gradient approximation
Glu	Glutamate
HA	Harmonic approximation
HF	Hartree-Fock
His	Histidine
HMO	Huckel molecular orbital
IR	Infrared
KS	Kohn-Sham
LCAO	Linear combination of atomic orbitals
LDA	Local-density approximation
MD	Molecular dynamics
MM	Molecular mechanics
MMP	Matrix metalloproteinase
MNDO	Modified neglect of diatomic overlap
NDDO	Neglect of diatomic differential overlap
PDB	Protein databank
PM3	Parametric method 3
QM	Quantum mechanics
RM1	Recife Model 1

SE	Semiempirical
Ser	Serine
Tyr	Tyrosine
XC	Exchange-correlation

1 Introduction

1.1 Water

Water is essential for human life and is arguably one of the most important molecules in physical, chemical, and biological systems. It performs different roles in living organisms, such as facilitating many chemical and biological reactions and participating to determine the structure and dynamics of proteins.[1,2] Many biochemical reactions occur in liquid water. For instance, water is good at solvating different ions and participates in proton transfer reactions; it works as a temperature buffer due to its high specific heat capacity, and it is a metabolite in photosynthesis, digestion, and aerobic respiration. In many cases, water actively participates in the functions of protein, mainly by forming hydrogen bonds and sometimes accepting/donating a proton from/to the protein.[3] Water has many anomalous properties including a complex phase diagram, a high dielectric constant, and a maximum density at 4°C.[2,4] Water is one of the most experimentally and theoretically studied liquids, and is still a fascinating topic for many researchers.[5–8]

Many of water's anomalies are thought to be the result of hydrogen bonding.[9,10] Hydrogen bonds and their orientation play crucial roles in determining the overall structure of water in its liquid and solid forms. Water retains its structural properties through hydrogen bonding.[11] Studying the diffusivity of water is very important to understand the transport related properties (e.g. water mass diffusion, hydrogen network relaxation, etc.) and these properties depend on the local structure and hydrogen bonding.[4] Studying the dielectric properties help to understand the dielectric behaviour of water.[12]

1.2 Description of H^+ and OH^- in liquid water

Protons (H^+) and hydroxide ions (OH^-) have been extensively studied in order to understand proton transport mechanisms in aqueous media, ranging from acid-base reactions to enzymatic proton transfer.[13–15] These ions also actively participate in catalytic reactions in biological systems.[16] Forming and losing strong covalent bonds between oxygen and hydrogen of

water molecule and weak hydrogen bonding between hydrogen and oxygen of inter water molecules in liquid create series of dynamically interconvertible hydrogen-bonded structures.[17]

Water often participates in chemical reactions by first decomposing into the more reactive species H^+ and OH^-. The tendency of the molecules to donate protons is measured by the acid dissociation constant. The acid dissociation constant, K_a, is the equilibrium constant of the reaction, which measures the strength of an acid in the solution. For the acid dissociation reaction, $AH \rightleftharpoons A^- + H^+$, K_a is given in terms of concentrations of conjugate base $[A^-]$, proton $[H^+]$, and acid [AH] as $K_a = [A^-][H^+]/[AH]$. K_a values are often expressed in terms of pK_a, which is defined as the negative logarithm of K_a ($pK_a = -\log K_a$). A lower value of pK_a means the compound is more acidic and has a higher tendency to release a proton. At standard ambient temperature and pressure conditions, the pK_a of water is 14.[18,19] Several factors play important roles in the decrease or increase of pK_as of compounds in solution such as resonance, atomic radius, and electronegativity of atoms.[20] Metal ions lower the pK_a of bound water and enhance its acidity.

Proton transport has crucial importance in a wide range of areas, from energy technologies (e.g. hydrogen fuel cells) to biological proton transport (e.g. H^+-ATPase). Proton transfer reactions also play crucial roles in energy conversion processes such as photosynthesis and cellular respiration.[21] The proton transport mechanism in water was first proposed in the early 18th century by von Grotthuss.[22] He hypothesized a mechanism that would later on be understood as a transfer through a network of hydrogen bond by forming hydronium ions and water molecules. Several studies have been done to elucidate the mechanism of proton transfer in liquid water by including the structure of the hydronium ion using multi-scale empirical valance-bond methodology.[13,23–25]

1.3 Description of hydrated metal ions

Metal hydration is defined as the transfer of a metal ion from the gas phase into water; for metal ion M^{n+}, it is simply $M^{n+}(g) \rightarrow M^{n+}(aq)$. In aqueous solution, metal-water binding mainly depends on the nature of the metal ion and ion-dipole interactions.[26] Since M–O bonds are – according to Pauling's first rule[27] – electrostatic in nature, the coordination number of hydrated

metal ions can be estimated from the ratio of the metal ionic radius to the radius of the oxygen atom of water. The ionic radius of a water ligand is 1.34Å.[28] The r_M/r_O ratios (r_M and r_O are ionic radii of the metal ion and oxygen in water respectively), expected coordination numbers, and configurations are presented in Table 1.1. The range of the r_M/r_O ratio estimates the coordination number and the corresponding configuration. The basic, expected coordination structure models are given in Figure 1.1. Due to specific electronic structures, a strong ability to form covalent interactions or a strong tendency to form cationic oxometal ions, some metal ions do not form regular hydration structures.

Table 1.1: r_M/r_O ratio, coordination number and configuration of metal ions.[27]

r_M/r_O	Coordination number	Configuration
0.000-0.155	2	Linear
0.155-0.225	3	Trigonal Planar
0.225-0.414	4	Tetrahedral
0.414-0.732	4	Square Planar
0.414-0.732	6	Octahedral
0.732-1.000	8	Square antiprism

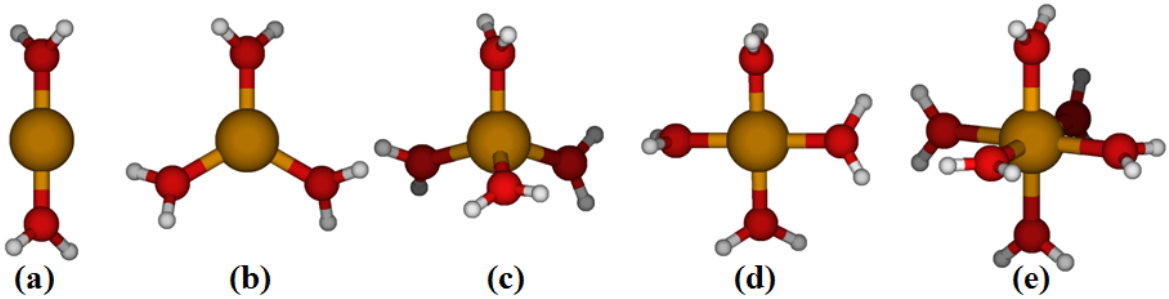


Figure 1.1: Basic hydration structures of a metal ion: (a) linear, (b) trigonal planar, (c) tetrahedral, (d) square planar, and (e) octahedral

Experimental[29] and theoretical[29,30] results show that Li^+ ions bind four water molecules in aqueous solution. Theoretical studies show that Na^+ ions have a tendency to bind 5-6 water molecules, whereas K^+ ions bind 5-8 water molecules in their first shell.[31,32] Divalent alkaline earth metals have clear, well-defined hydration structures, with Be^{2+} being tetrahedral, Mg^{2+} octahedral, and Ca^{2+} square antiprism with coordination numbers 4, 6, and 8 respectively.[33–36] Experimental studies show that both Zn^{2+} [34,35] and Fe^{2+} [37] are hexa-coordinated in aqueous solution.

Structures of metal-ligand complexes depend on the metal and its oxidation state.[38] The flexibility of a metal ion's coordination is one of the important factors for its catalytic roles in biological systems.[39,40] The energy penalties for changing the coordination environment of divalent metal ions relative to their minimum energy coordination have been studied.[41,42] The relative energies of hydrated metal ion complexes obtained from ab initio molecular orbital calculations (MP2 level of theory) for different coordination numbers are presented in Table 1.2. Energy penalties for changing coordination numbers of Be^{2+} and Mg^{2+} are high. $[Be(H_2O)_4]^{2+}\cdot 2H_2O$ is 22.0 kcal/mol more stable than $[Be(H_2O)_6]^{2+}$, and Be^{2+} cannot form a stable penta-coordinated structure in its inner sphere. In the case of magnesium, $[Mg(H_2O)_6]^{2+}$ is 9.0 and 4.0 kcal/mol more stable than $[Mg(H_2O)_4]^{2+}\cdot 2H_2O$ and $[Mg(H_2O)_5]^{2+}\cdot 1H_2O$, respectively. Due to the high energy penalties for Be^{2+} and Mg^{2+}, their inner coordination spheres are fairly rigid. By contrast, Ca^{2+} and Zn^{2+} have more variable coordination numbers with very low energy penalties except in the case of $[Ca(H_2O)_6]^{2+}$. Zn^{2+} has lower energy penalties than Ca^{2+} for changing coordination from higher (hexa) to lower (penta). Thus, Zn^{2+} is involved in a wide range of chemical reactions including reactions for which a change in coordination number is necessary.[43]

Table 1.2: Relative energies in kcal/mol of hydrated metal ion complexes. *m* and *n* are the numbers of water molecules in the first and second coordination shells, respectively. (Data are taken from references [41] and [42])

Metal ion	Energy	[*m*,*n*]	Energy	[*m*,*n*]	Energy	[*m*,*n*]
Be^{2+}	0.0	[4,2]	+22.0	[6,0]		
Mg^{2+}	0.0	[6,0]	+4.0	[5,1]	+9.0	[4,2]
Ca^{2+}	0.0	[6,0]	+8.2	[5,1]	+15.0	[4,2]
	0.0	[7,0]	–1.4	[6,1]	+4.0	[5,2]
	0.0	[8,0]	+1.0	[7,1]	–0.5	[6,2]
Zn^{2+}	0.0	[6,0]	+1.0	[5,1]	+1.4	[4,2]
	0.0	[5,0]	+0.6	[4,1]		
	0.0	[4,0]	+14.6	[3,1]		

1.4 Preferential coordination of metal ions

Metal ions do not only activate or regulate the function of proteins by causing a change in the structure and/or serving as catalytic centers for chemical reactions, but they are also involved in protein stabilization, protein-protein interactions, and protein-ligand binding.[44] Analyses of the local coordination environment of metal ions for their binding preferences with different types of ligands mainly containing oxygen, nitrogen and sulfur are discussed for Mg^{2+}, Ca^{2+}, and Zn^{2+}. Cambridge Structural Database (CSD) analysis shows that both Mg^{2+} and Ca^{2+} have a strong tendency to bind with oxygen rather than nitrogen and sulfur.[42] Zinc, however, can easily bind nitrogen and sulfur ligands (in addition to oxygen ligands). The binding preference of zinc for oxygen, nitrogen, or sulfur ligands depends on the coordination structure it adopts. For example, zinc tends to bind with nitrogen ligands in tetrahedral structures, whereas it favors the binding of oxygen ligands in octahedral structures.[42]

Table 1.3: Occurrence of ligands for a given coordination around Zn^{2+}, Mg^{2+}, and Ca^{2+} ions in protein structures. (Data from ref. [45])

Ligands/ coordination	Zn		Mg			Ca			
	3	4	3	4	5	3	4	5	6
Cys (C)	24%	56%	-	-	-	-	-	-	-
His (H)	37%	26%	5%	3%	3%	1%	-	-	-
Asp/Glu (D/E)	25%	9%	37%	39%	22%	42%	42%	43%	43%
Backbone Oxygen	3%	2%	28%	31%	41%	34%	34%	33%	34%
Asn/Gln (N/Q)	2%	1%	5%	6%	9%	7%	8%	10%	10%
Other	9%	6%	25%	21%	25%	16%	16%	14%	13%

Table 1.3 lists the occurrence of different amino acid ligands around the metal ions Zn^{2+}, Mg^{2+}, and Ca^{2+} for a given coordination configuration.[45] In a protein, a zinc ion normally binds with three or four amino acids, whereas Mg^{2+} binds with three to five amino acids and Ca^{2+} binds with three to six amino acids. In line with the CSD analysis, Mg^{2+} and Ca^{2+} do not bind with cysteine at all and bind with histidine in a small fraction of total proteins. They prefer to associate with aspartic or glutamic amino acids and backbone oxygens. Zn^{2+}, on the other hand, binds with both cysteine and histidine. Unlike Ca^{2+} and Mg^{2+}, Zn^{2+} does not prefer to bind with backbone oxygen. Most of the time, zinc binds ligands in a tetrahedral configuration, whereas Mg^{2+} and Ca^{2+} bind in several configurations equally. The coordination numbers of metal ions decrease as size of ligands increases. Consequently, the coordination number of metal ions in proteins is lower than in water, because amino acid molecules have a larger size than water molecules.

The selectivity of metal ions towards biological donor ligands is very important for their specific roles. Pearson[46,47] classified metal ions as 'hard' or 'soft' based on their size and polarizability. Hard metal ions (or hard acids) have a high positive charge, small size, and very low polarizability. Soft metal ions (or soft acids) have low positive charge, large size, and high polarizability. Pearson's hard-soft acid-base theory[46,47] enables the empirical ordering of metal

ions according to their preferences for organic ligands. Simply, it explains that hard acids tend to bind to hard bases, whereas soft acids tend to bind to soft bases. A recent study shows that hard-soft acid-base theory can predict the selectivity of metal ions for biological donor ligands.[48] Being hard acids, Na^+, K^+, Mg^{2+}, and Ca^{2+} tend to bind to hard bases such as carbonate, sulphate, and carboxylate ligands, whereas soft acids Cu^+, Au^+, and Ag^+ favour binding to soft bases such as phenyl, thiol, and cyanide groups. Borderline acids Cu^{2+}, Zn^{2+}, Co^{2+}, and Fe^{2+} bind to aniline, imidazole, pyridine, nitrite, etc.

1.5 Metal ions in biological systems

Metal ions actively participate in approximately one-third of all enzymes and they have numerous applications in biological processes.[44,49–52] Na^+ and K^+ ions participate in controlling blood pressure, propagation of nerve impulses, and act as counter ions for DNA and membranes.[53–55] Iron containing hemoglobin carries oxygen in blood,[49] copper-containing proteins (e.g. plastocyanin) are involved in electron transport process during photosynthesis.[56] Metals are classified into three categories based on the amount in which they present in biological systems, namely bulk, trace, and ultra-trace. Sodium, magnesium, potassium, and calcium ions are abundant in biological systems, and they are considered as bulk metals. Iron, zinc and copper are trace metals, and barium, bromine, and scandium are ultra-trace metals. Despite the small amounts present in biological systems, these trace and ultra-trace ions play very significant functions.[44]

1.6 Zinc metalloproteins

Zinc is one of the most important metals in biological systems serving various functions performed by zinc metalloproteins. A zinc deficiency causes many health problems such as depression, Alzheimer's and Parkinson's diseases, diabetes, etc.[57] In addition, a zinc deficiency can lead in serious impact on growth and on immune systems.[58] Zinc binding sites are mainly classified into structural, catalytic, and co-catalytic.[59,60] Examples of the structural and catalytic sites in zinc metalloproteins are shown in Figure 1.2. Generally, zinc binds four amino acids at structural sites and three amino acids and one water molecule at catalytic sites. Zinc also bridges with second metal ion (normally with either zinc or copper) for co-catalytic functions.[60]

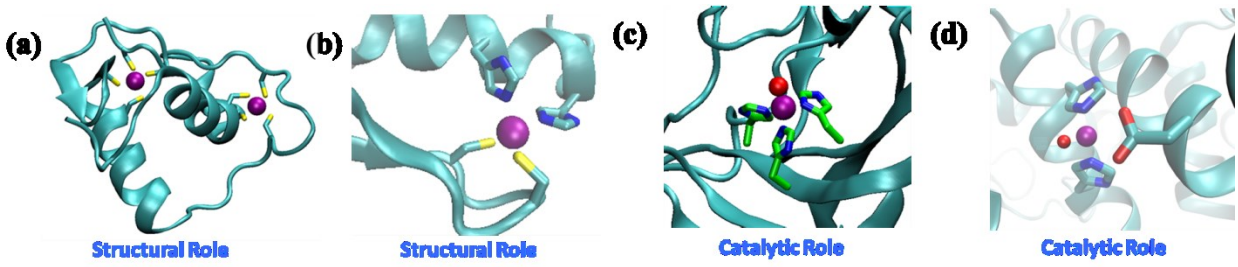
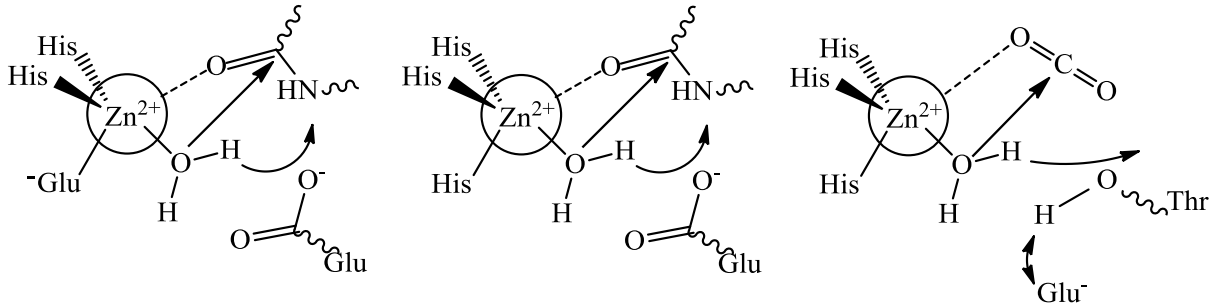


Figure 1.2: Examples of structural and catalytic sites of zinc metalloproteins. The Zn^{2+} active site is coordinated in (a) with four cysteines in breast cancer-associated protein 2 (PDB code: 2DGA), in (b) with two histidines and two cysteines in DNA-binding zinc finger (PDB code: 1A1J), in (c) with three histidines and one water molecule in carbonic anhydrase (PDB code: 1CA2), and in (d) with two histidines, one glutamate, and one water molecule in thermolysin (PDB code: 1LNF).

Generally in zinc metalloenzymes, the zinc cation is coordinated tetrahedrally with three amino acids and one water molecule. In this study, the aim is to develop theoretical models so that they can be applied to zinc metalloproteins to study various properties. Our target zinc enzymes are zinc hydrolases, mainly thermolysin, matrix metalloproteinases (MMPs), and carbonic anhydrases (CAs). These enzymes are widely studied, and they share similar binding site motifs. Reaction centers for all three classes are shown in Figure 1.3. The zinc ion mainly has two roles; first, it facilitates zinc-bound water to deprotonate, which allows zinc-bound hydroxyl to be involved in a nucleophilic attack of the substrate (which is carbon dioxide in the case of carbonic anhydrases), and second, zinc polarizes the carboxyl group involved in the peptide bond of the substrate.



Thermolysin Matrix Metalloprotease Carbonic Anhydrase

Figure 1.3: The reaction center of Thermolysin, Matrix Metalloprotease, and Carbonic Anhydrase. Arrows indicate the physical motion of hydrogen and oxygen atoms.

Thermolysin is a thermostable zinc enzyme. It serves important catalytic roles in the hydrolysis of peptide bonds. Thermolysin is an endopeptidase, an enzyme that catalyzes hydrolysis of peptide bonds. At the active site of thermolysin, the zinc cation forms a distorted tetrahedral configuration with three amino acid side chains (two histidines and one glutamate) and one water molecule.[61] For catalytic roles, one water molecule is always present at the active site retaining the tetrahedral configuration.[62]

Matrix metalloproteinases (MMPs) are involved in degradation of extracellular components and also have significant roles in cancer as markers.[63] These enzymes also serve important roles in embryonic development, morphogenesis, reproductive tissue resorption, and remodelling/repair tissues.[64] They are zinc endopeptidases as well. In MMP, peptide cleavage is realized through the following steps: proton transfer from zinc-bound water with the help of a neighbouring glutamate, following the nucleophilic attack on the carboxyl group of the peptide by the hydroxyl group bound to the zinc cation. From previous experiments, it has been confirmed that glutamate plays an important role in maintaining catalytic activity.[65]

Carbonic anhydrase is a zinc enzyme, which aids in the conversion of carbon dioxide to bicarbonate and vice versa. In carbonic anhydrase, zinc facilitates zinc-bound water to deprotonate; zinc-bound hydroxide ion participates in the nucleophilic attack to the carbon of carbon dioxide, and converts carbon dioxide into bicarbonate.

1.7 Motivation and overview of the thesis

Metal ions are involved in protonation/deprotonation of bound ligands and they stabilize clusters of negatively charged ligands.[66] They participate in important local interactions resulting in global changes in protein structure such as metal-coupled protein folding/unfolding.[66] As many of these interactions involve electrons motion, electronic structure theory methods are needed for accurate description of the motion of electrons in atoms and molecules. High-level quantum mechanical methods are computationally very expensive and only apply to systems with a few tens of atoms for molecular dynamic simulations of a few picoseconds. Picosecond time-scale is not long enough to see atom- or molecule-scale motions such as proton transfer, side chain flip, solute permeation.[67,68]

Current state-of-the-art for molecular modelling of chemical reactions in proteins is the hybrid quantum mechanical/molecular mechanical (QM/MM) approach, in which the active site molecules (including metal-binding sites and molecules participating in the chemical reactions) are treated at a conventional DFT level and remaining atoms of the protein and the solvent are described using classical force fields. These techniques still require a significant amount of computing power and only effectively apply to molecular dynamics simulations for up to a few hundred picoseconds. In terms of computational cost, the most expensive part in the Hartree-Fock based methods is to calculate and store the two-electron integrals.[69] Semi-empirical (SE) quantum methods approximate many of these integrals by introducing empirical parameters to represent the approximations. These parameters are obtained by fitting empirical data obtained from experiments and/or high-level ab initio calculations. Because of the approximations, the computational cost of SE methods is reduced by more than three orders of magnitude when compared with a conventional DFT calculation. Hence, systems with a large number of atoms can be modelled using SE methods. SE methods can be used to describe the full active site of the protein including first and second shell amino acids. Using semiempirical models, we can easily simulate a system of a few hundred atoms for molecular dynamics simulations of a few nanoseconds that allows us to see proton transfer, amino acid side-chains relaxation, etc. Simulations on the nanosecond time-scale also allow examining how water molecules enter and exit the active site of the proteins. Because semiempirical methods have simplified mathematical formulations, one can easily develop reaction-specific[70,71] SE models to reduce transferability issues and increase the performance of existing models.

The main focus of this thesis is to develop new SE models based on neglect of diatomic differential overlap (NDDO)[72] approximations using the AM1 (*Austin model 1*)[73] type Hamiltonian. These models are re-parameterized to reproduce the important structural and chemical properties of bioinorganic clusters containing zinc metal and biological molecules displaying zinc–binding motifs analogous to zinc metalloproteins. The models are used to reproduce the correct molecular structures and reaction energies of the metal-containing clusters. Various crucial properties (e.g. proton affinity, proton transfer, structural dynamics, hydrogen bonding, reaction and activation energies, and pK_a values) will be analyzed in different systems (e.g. gas-phase clusters, liquid water, and metal solvated aqueous solutions).

The work is divided into three main parts. First, AM1 parameters for hydrogen and oxygen are re-parameterized to describe proton transfer reactions in gas-phase as well as the static and dynamic properties of liquid water. Second, using previously optimized parameters for H and O, AM1 parameters for sulfur and zinc are re-parameterized to study zinc-water and zinc-cysteine binding in zinc metalloproteins. Third, using parameters for H, O, S, and Zn, re-parameterization of AM1 model for carbon and nitrogen is performed to study nucleophilic attacks on the backbone, peptide hydrolysis, and proton transfer reactions in zinc metalloproteases. The overall development workflow of SE models is shown in Figure 1.4.

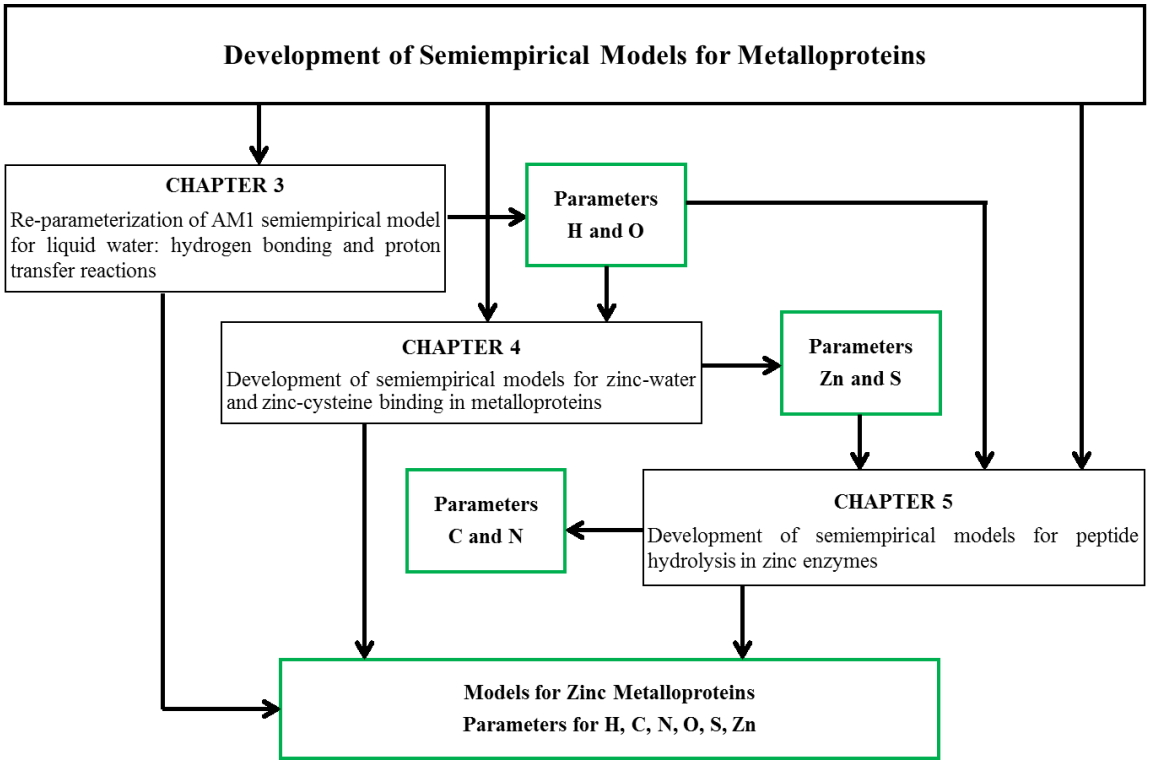


Figure 1.4: Workflow for the development of semiempirical models for zinc metalloproteins.

The thesis is divided into six chapters. Chapter 2 describes the methodologies, where mathematical formulations of SE models are briefly discussed. Genetic algorithms for SE parameterization, formulation of molecular dynamics simulations, and mathematical expressions for calculation of infrared spectra and pK_a for different systems are also presented. Chapter 3 presents the re-parameterization of the AM1 model for proton transfer reactions in water and validation of the models to describe the static and dynamic properties of liquid water. Chapter 4 describes the re-parameterization of the AM1 model for describing various properties, including metal-induced pK_a shifts, in zinc-water, and zinc-cysteine binding in zinc enzymes. Chapter 5 presents the re-parameterization of SE models for peptide hydrolysis in zinc metalloproteases. Finally, Chapter 6 highlights the conclusions of the thesis and presents an outlook for future research.

2 Methods

2.1 General formulation of quantum mechanical methods

The time-independent, nonrelativistic Schrödinger equation is written as

$$\hat{H}\Psi = E\Psi \tag{2-1}$$

where $\hat{H}$, E, and Ψ are Hamiltonian operator, energy eigenvalue and wavefunction (eigenfunction) of the system, respectively.

The Hamiltonian for M nuclei and N electrons can be written as[74]

$$\hat{H} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{R_{Ai}} + \sum_{i=1}^{N}\sum_{j>1}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \tag{2-2}$$

where i and j indicate the N electrons and A and B indicate the M nuclei in the system. The first two terms of eq. (2-2) are the kinetic energy of the electrons and nuclei A. The remaining terms give the potential energy. The first is a nucleus-electron attraction, the second electron-electron repulsion, and the third nucleus-nucleus repulsion energy term. M_A is the mass of the nucleus of atom A, and Z_A and Z_B are the nuclear charges of atoms A and B. R_{AB} is the distance between atoms A and B and r_{ij} is the distance between the i^{th} and j^{th} electrons. ∇^2 is the Laplacian operator.

The mass of the nucleus is much larger than that of electron, and the nuclei move much more slowly than the electrons. It is considered that electrons move in the field of fixed nuclei. All nuclear degrees of freedom (and the energy terms related) are treated as constant on the time scale of the electronic motion. Born-Oppenheimer approximation is the assumption that electronic and nuclear motion in the system can be separated. Using this approximation, we can separate total electronic and nuclear parts from eq. (2-2). As a result,

$$\begin{aligned}\hat{H}_{\text{ele}} &= -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{R_{Ai}} + \sum_{i=1}^{N}\sum_{j>1}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \\ &= \hat{T} + \hat{V}_{\text{Ne}} + \hat{V}_{\text{ee}}\end{aligned} \tag{2-3}$$

and constant nuclear repulsion term, electronic term , and total energy of the system become

$$E_{\text{nuc}} = \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}} \tag{2-4}$$

$$\widehat{H}_{\text{ele}} \Psi_{\text{ele}} = E_{\text{ele}} \Psi_{\text{ele}} \tag{2-5}$$

$$E_{\text{tot}} = E_{\text{ele}} + E_{\text{nuc}} = E_{\text{ele}} + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}} \tag{2-6}$$

The electronic wavefunction Ψ_{ele} is a function of electron coordinates and the spin of N electrons of the system.

$$\Psi_{\text{ele}} = \Psi(r_1, r_2, \dots\dots\dots, r_N; \omega_1, \omega_2, \dots\dots\dots, \omega_N) \tag{2-7}$$

where r_i is the position vector of i^{th} electron and ω_i is its spin.

According to Hartree scheme[74], the electronic wavefunction can be written as the product of N one-electron wavefunctions. As a result, the electronic Hamiltonian operator $\widehat{H}_{\text{ele}}$ and electronic energy E_{ele} reduced to

$$\widehat{H}_{\text{ele}} \approx \sum_{i=1}^{N} \hat{h}_i \quad \text{and} \quad E_{\text{ele}} \approx \sum_{i=1}^{N} \varepsilon_i \tag{2-8}$$

where $\hat{h}_i$ is the one-electron Hamiltonian of i^{th} electron and ε_i is corresponding energy.

However, since the electron has nonzero spin, it must obey the Pauli's exclusion principle. According to Pauli's exclusion principle, wavefunction must be antisymmetric with respect to exchanging any two electrons. Eq. (2-5) is only exactly solvable for one electron system and hence, it cannot apply for N-electron system. Hartree-Fock approximation allows us to estimate many particle solutions of N-electron wavefunction by antisymmetrized product of N one-electron wavefunction $\chi_i(\vec{\mathrm{x}}_i)$. This product is also called Slater determinants Φ_{SD}.

$$\Phi_{\text{SD}} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\vec{\mathrm{x}}_1) & \chi_2(\vec{\mathrm{x}}_1) & \cdots & \chi_N(\vec{\mathrm{x}}_1) \\ \chi_1(\vec{\mathrm{x}}_2) & \chi_2(\vec{\mathrm{x}}_2) & & \chi_N(\vec{\mathrm{x}}_2) \\ \vdots & \vdots & & \vdots \\ \chi_1(\vec{\mathrm{x}}_N) & \chi_2(\vec{\mathrm{x}}_N) & & \chi_N(\vec{\mathrm{x}}_N) \end{vmatrix} \tag{2-9}$$

The variational principle[74] states that the ground state energy, E_0, is always equal or less than the expectation value of the Hamiltonian, $\widehat{H}$, with the trial wavefunction, ψ_{trial}. i.e.

$$E_0 \leq \langle \psi_{\text{trial}} | \hat{H} | \psi_{\text{trial}} \rangle \tag{2-10}$$

Thus, the minimum energy (energy of the ground state) is calculated by varying the trial wavefunction until the expectation value of the Hamiltonian is minimized. By invoking the variational principle, the *best* Slater determinants (the particular Φ_{SD}, which yields the lowest energy) and Hartree-Fock (HF) energy can be found. The HF energy can be written as[74]

$$E_{\text{HF}} = \langle \Phi_{\text{SD}} | \hat{H} | \Phi_{\text{SD}} \rangle = \sum_i^N \langle i | \hat{h} | i \rangle + \frac{1}{2} \sum_i^N \sum_j^N \langle ij | ij \rangle - \langle ij | ji \rangle \tag{2-11}$$

where,

$$\langle i | \hat{h} | i \rangle = \int \chi_i^*(\vec{\mathrm{x}}_1) \left\{ -\frac{1}{2} \nabla^2 - \sum_A^M \frac{Z_A}{r_{1A}} \right\} \chi_i(\vec{\mathrm{x}}_1) d\vec{\mathrm{x}}_1 \tag{2-12}$$

$$\langle ij | ij \rangle = \iint |\chi_i(\vec{\mathrm{x}}_1)|^2 \frac{1}{r_{12}} |\chi_j(\vec{\mathrm{x}}_2)|^2 d\vec{\mathrm{x}}_1 d\vec{\mathrm{x}}_2 \tag{2-13}$$

$$\langle ij | ji \rangle = \iint \chi_i(\vec{\mathrm{x}}_1) \chi_j^*(\vec{\mathrm{x}}_1) \frac{1}{r_{12}} \chi_j(\vec{\mathrm{x}}_2) \chi_i^*(\vec{\mathrm{x}}_2) d\vec{\mathrm{x}}_1 d\vec{\mathrm{x}}_2 \tag{2-14}$$

Eq. (2-12), (2-13), and (2-14) are the contributions due to the kinetic energy and electron-nucleus attraction, Coulomb repulsion of electrons, and Hartree-Fock exchange energy resulting from quantum (fermion) nature of the electrons. In general, Hartree-Fock method does not count instantaneous electron-electron repulsion correctly, which results errors in the wavefunction and the correlation energy. Another problem of HF approximation is the definition of *spin orbitals*. To resolve this issue, a linear combination of atomic orbital (LCAO) approximation is normally used where molecular orbitals are written as the linear combination of atomic orbitals.

$$\chi_i = \sum_j^{N_{\text{AOs}}} C_j^i \phi_j \tag{2-15}$$

where N_{AOs} are the number of atomic orbitals of the system, ϕ_j is the atomic orbital in the molecular orbital χ_i, and C_j^i is the coefficient of atomic orbital ϕ_j.

2.2 Density functional theory

DFT is a popular quantum mechanical methods used to compute the ground state electronic structure of many-body systems. DFT techniques are used for atoms, molecules, and solids to calculate different properties including molecular structure, electronic and magnetic properties, optical and vibrational properties, reaction paths etc.[74] In DFT method, properties of many-electron systems can be evaluated using a functional which in this case is the spatially dependent electron density. Type of the DFT method, accuracy, and computational cost depend on the type of functional used. Some of the popular functional forms are local density approximation (LDA), generalized gradient approximation (GGA), and hybrid density functional techniques. In LDA, exchange and correlation functionals only contain terms related to electron density. LDA functional is computationally very fast to calculate. LDA performs well for systems where density varies slowly in space whereas, for strongly correlated systems, the LDA is very inaccurate.[75] LDA does not account for van der Waal bonding and gives very inaccurate hydrogen bonding.[76] Unlike LDA, GGA functionals use both value of electron density and its gradient. Comparative study of LDA and GGA shows that GGA gives remarkable accurate results compared to LDA for some properties of molecules including hydrogen bonding.[77] For small to moderate-sized chemical systems, hybrid density functional techniques are often used. In hybrid methods, the exchange function is often written as a linear combination of the Hartree-Fock exchange (eq. (2-14)) and a function (which is a function of electron density and its gradient). B3LYP[78] is a widely used hybrid DFT method based on Becke's 3-parameters functional (B3)[79] with Lee, Yang, and Parr's (LYP) nonlocal correlation.[80]

In Kohn-Sham DFT formulation[81], the exchange term in eq. (2-11) is replaced by exchange-correlation functionals $E_{\mathrm{XC}}[\rho]$, which counts both exchange and correlation energies with exchange $E_{\mathrm{X}}[\rho]$ and correlation $E_{\mathrm{C}}[\rho]$ functionals. Then the final simplified form for Kohn-Sham energy is[74]

$$E_{\mathrm{KS}} = \sum_{i}^{N} \langle i | \hat{h} | i \rangle + \frac{1}{2} \sum_{i}^{N} \sum_{j}^{N} \langle ij | ij \rangle + E_{\mathrm{XC}}[\rho] + E_{\mathrm{nuc}} \tag{2-16}$$

$$E_{\mathrm{XC}}[\rho] = E_{\mathrm{X}}[\rho] + E_{\mathrm{C}}[\rho] \tag{2-17}$$

$$E_{\mathrm{KS}} = \sum_{i}^{N} \langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{i}^{N}\sum_{j}^{N} \langle ij|ij\rangle + E_{\mathrm{X}}[\rho] + E_{\mathrm{C}}[\rho] + E_{\mathrm{nuc}} \qquad (2\text{-}18)$$

Hartree-Fock theory is a special case of Kohn-Sham density functional theory with $E_{\mathrm{X}}[\rho]$ given by the exchange integral (eq. (2-14)) and $E_{\mathrm{C}}[\rho] = 0$. DFT method depends on the type of exchange and correlation functionals used. Explicitly,

$$E_{\mathrm{XC}}^{\mathrm{LDA}}[\rho] = \int \rho(\vec{\mathrm{r}})\, \varepsilon_{\mathrm{XC}}\, \rho(\vec{\mathrm{r}})\, \mathrm{d}\vec{\mathrm{r}} \qquad (2\text{-}19)$$

$$E_{\mathrm{XC}}^{\mathrm{GGA}}[\rho_\alpha, \rho_\beta] = \int \mathrm{f}(\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta)\, \mathrm{d}\vec{\mathrm{r}} \qquad (2\text{-}20)$$

$$E_{\mathrm{XC}}^{\mathrm{B3LYP}} = (1-\mathrm{a})E_{\mathrm{X}}^{\mathrm{LDA}} + \mathrm{a}E_{\mathrm{XC}}^{\lambda=0} + \mathrm{b}E_{\mathrm{X}}^{\mathrm{B88}} + \mathrm{c}E_{\mathrm{C}}^{\mathrm{LLYP}} + (1-\mathrm{c})E_{\mathrm{C}}^{\mathrm{LDA}} \qquad (2\text{-}21)$$

2.3 Semiempirical molecular orbital theory based on NDDO approximation

Neglect of diatomic differential overlap (NDDO)[72] methods rely on two main basic approximations.

1. For closed and open-shell molecules, semiempirical formulation uses frozen core approximation, which means that the valence electrons move in the field of a fixed core composed of the nuclei and inner shell electrons. The valence shell molecular orbitals χ_i can be represented as a linear combination of atomic orbitals ϕ_ν,

$$\chi_i = \sum_{\nu} C_{\nu i}\phi_\nu \qquad (2\text{-}22)$$

2. Overlap matrix $S_{\mu\nu} = \int \phi_\mu(1)\phi_\nu(1)d\tau_1$[82] is unity if both atomic orbitals are equal otherwise null. Thus all the two centered overlapping integrals between orbitals on different atoms are neglected. All three centered and four centered two-electron integrals are completely neglected.

The LCAO applications to Hartree-Fock approximation[82] gives the secular Roothaan-Hall equation[83,84]

$$\sum_{\nu}(F_{\mu\nu} - S_{\mu\nu}E_i)C_{\nu i} = 0 \qquad (2\text{-}23)$$

where $F_{\mu\nu}$ and E_i are Fock matrix and energy of corresponding molecular orbital.[82] Coefficients $C_{\nu i}$ are evaluated using eq. (2-23). Fock matrix approximates the single-electron energy operator in a given set of basis functions.

Following the NDDO approximation, overlap matrix $S_{\mu\nu}$ is replaced with Kronecker-delta $\delta_{\mu\nu}$ which is 1 if $\mu = \nu$ and 0 otherwise. Then eq. (2-23) becomes,

$$\sum_{\nu}(F_{\mu\nu} - \delta_{\mu\nu}E_i)C_{\nu i} = 0 \tag{2-24}$$

For atomic orbitals ϕ_μ and ϕ_ν centered on atom A and atomic orbitals ϕ_λ and ϕ_σ centered on atom B (≠A), the matrix elements $F_{\mu\nu}$ can be written as:[72]

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{B}\sum_{\lambda,\sigma}^{B} P_{\lambda\sigma}(\mu\nu|\lambda\sigma) - \frac{1}{2}\sum_{\lambda,\sigma}^{B} P_{\lambda\sigma}(\mu\sigma|\nu\lambda) \qquad \mu,\nu \text{ both on } A \tag{2-25}$$

$$F_{\mu\nu} = H_{\mu\nu} - \frac{1}{2}\sum_{\sigma}^{A}\sum_{\lambda}^{B} P_{\lambda\sigma}(\mu\sigma|\nu\lambda) \qquad \mu \text{ on } A, \nu \text{ on } B \tag{2-26}$$

Density matrix $P_{\lambda\sigma}$, two-electron integrals $(\mu\nu|\lambda\sigma)$ and core Hamiltonian $H_{\mu\nu}$ are given by[82]

$$P_{\lambda\sigma} = 2\sum_{i}^{\text{occ}} C_{i\lambda}^{*}C_{i\sigma} \tag{2-27}$$

$$(\mu\nu|\lambda\sigma) = \iint \phi_\mu^*(1)\phi_\nu^*(1)\frac{1}{r_{12}}\phi_\lambda^*(2)\phi_\sigma^*(2)d\tau_1 d\tau_2 \tag{2-28}$$

$$H_{\mu\nu} = U_{\mu\nu} - \textstyle\sum_{B(\neq A)}\langle\mu|V_{\text{B}}|\nu\rangle \quad \phi_\mu, \phi_\nu \text{ are on A} \tag{2-29}$$

where $U_{\mu\nu}$ is one electron matrix element. Total electronic energy of the valence electrons is[82]

$$E_{\text{ele}} = \frac{1}{2}\sum_{\mu,\nu} P_{\mu\nu}(H_{\mu\nu} + F_{\mu\nu}) \tag{2-30}$$

The total energy of the molecule is the sum of electronic energy and core repulsion energy (E_{AB}^{core}) between the cores of A and B.

$$E_{\text{tot}}^{\text{mol}} = E_{\text{ele}} + \sum_{A}\sum_{A<B} E_{AB}^{\text{core}} \tag{2-31}$$

The NDDO-based semiempirical models differ from one another according to the core repulsion function (CRF) terms, the type of basis set functions, and the parameterization techniques. Two commonly used models are modified neglected diatomic overlap (MNDO) model[85] and the Austin model 1 (AM1)[73] for which the CRFs are defined as

$$E_{AB}^{\text{core}}(\text{MNDO}) = Z_A Z_B (s^A s^A, s^B s^B)[1 + \mathrm{e}^{-\alpha_A R_{AB}} + \mathrm{e}^{-\alpha_B R_{AB}}] \tag{2-32}$$

$$E_{AB}^{\text{core}}(\text{AM1}) = Z_A Z_B (s^A s^A, s^B s^B)[1 + F(A) + F(B)] \tag{2-33}$$

$$F(A) = \mathrm{e}^{-\alpha_A R_{AB}} + \sum_i K_{A,m} \mathrm{e}^{L_{A,m}\left(R_{AB} - M_{A,\mathrm{m}}\right)^2} \tag{2-34}$$

$$F(B) = \mathrm{e}^{-\alpha_B R_{AB}} + \sum_j K_{B,n} \mathrm{e}^{L_{B,n}\left(R_{AB} - M_{B,n}\right)^2} \tag{2-35}$$

Z_A and Z_B are core charges of atoms A and B, s^A and s^B are s-type atomic orbitals, α_A and α_B are core-core repulsion term for atom A and atom B, and R_{AB} is the internuclear distance between atoms A and B. K, L, and M are the amplitude, steepness, and displacement of Gaussian functions of atoms A and B. At most four Gaussian terms are used in AM1 with $m = 1,\ldots,4$ and $n = 1,\ldots,4$. These extra Gaussian terms differ AM1 from MNDO. PM3[86] is another widely used SE methods, which is also based on NDDO approximations. Instead of four (as in AM1), PM3 uses only two extra Gaussian terms in the core repulsion function.

The earliest SE method developed was the Hückel Molecular Orbital (HMO) method which was introduced to permit qualitative study of the purely π-electrons systems in 1930.[87] SE methods can predict correct structural, energetic and spectroscopic properties of molecular and solid-state systems. New SE methods are currently being developed for the applications to large complexes including proteins. Recent study shows that it is possible to use parameterized semiempirical methods (e.g. RM1_{BH}: RM1 for biological hydrogen bonding) to investigate the weak interaction of biological systems especially binding energy calculation of hydrogen bonds.[88] It is also shown that the addition of extra terms in the CRF improves the description of hydrogen bonds.[89–91] For the enzymes that catalyze redox reactions, especially for redox-active elements iron and copper in their active sites, the semiempirical methods MNDO/d[92] or AM1/d[93]

or modified semiempirical methods with *spd*-orbitals can be applied. The accuracy of semiempirical methods can be improved by modifying the core-repulsion function.

2.4 Genetic algorithms for parameterization

Genetic algorithms (GAs) are heuristic search techniques inspired from Darwin's theory of evolution and natural selection. GA first constructs a random initial population of chromosomes (i.e. the parameters of the SE models) and evaluates the fitness of each member. A new population is created by breeding. Breeding is done by taking two fit chromosomes from parents and intermingling to create new offspring. The fitness of each member in new population is evaluated and the old population is replaced partially or completely. During the reproduction stage, two chromosomes from the gene pool are selected and two new chromosomes are created by random mutation and crossover.

Crossover is the exchange of the genes between two parents to create the resultant offspring. It is done by swapping genes at randomly chosen positions (see Figure 2.1 (a)). During two-point crossover, two random splicing points are chosen along two parents string and string portion within selected splice points is exchanged. Crossover is a very powerful technique to have genetic diversity.

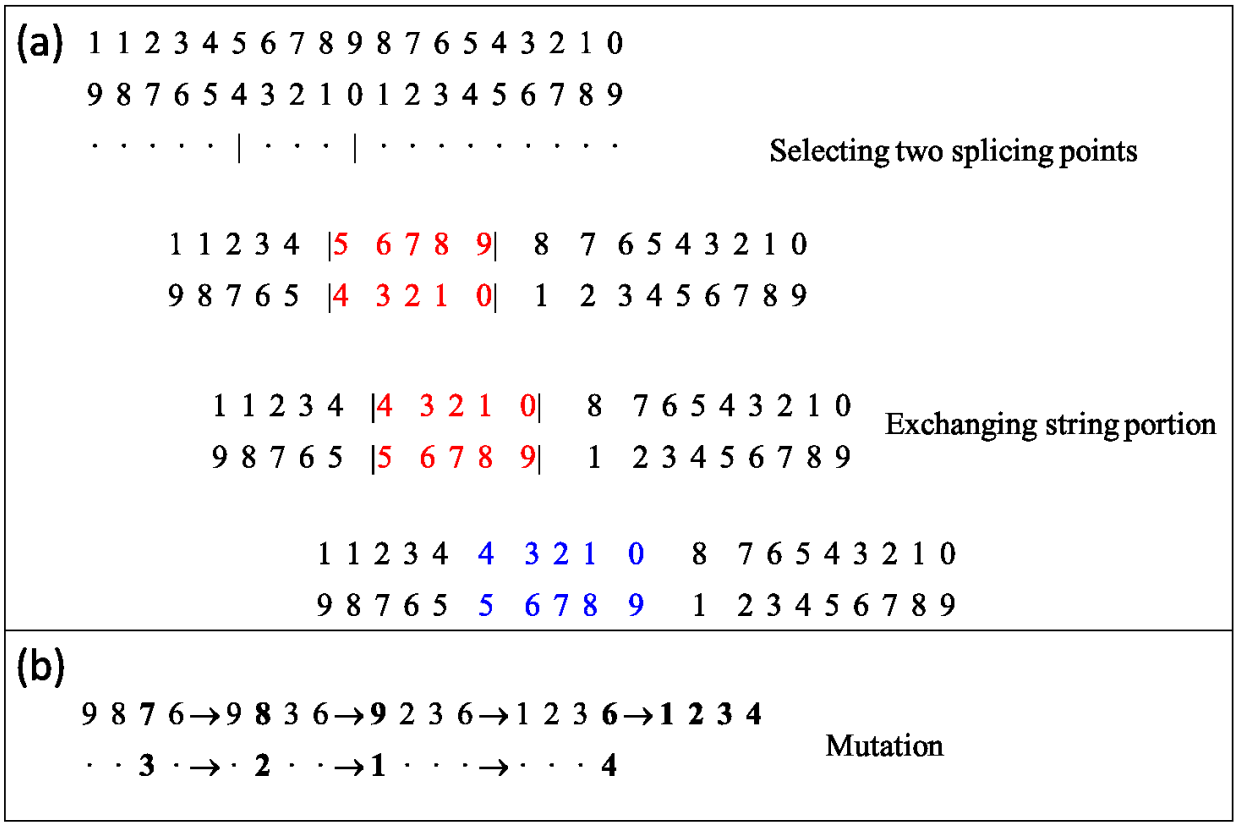


Figure 2.1: Examples of (a) two-point crossover and (b) mutation operations in genetic algorithms.

Mutation is carried out by changing random genes in an individual (see Figure 2.1 (b)). It is done one at a time and fitness score is calculated each time. Only sequences with higher fitness scores are collected, if the fitness score is low, another mutation is performed. During the optimization, mutation can also help to avoid local minima coming from the population with similar chromosomes.

In this work, a parallel version of PIKAIA,[94] a genetic algorithm for function optimization, is used to optimize the parameters of the SE models. PIKAIA does not calculate the derivatives of the "goodness-of-fit" function with respect to the parameters, and thus it is one of the most computationally efficient genetic algorithms. PIKAIA uses decimal encoding. Decimal encoding is better than binary encoding because binary encoding uses platform-dependent functions in FORTRAN. PIKAIA offers different parameters such as a number of individuals in a population, a number of generations over which the solution is evolved, and a number of significant digits.

2.5 Molecular dynamic simulations

Molecular dynamic (MD) is a powerful computer simulation technique in molecular modeling that enables to describe fluctuations of the molecular structures at the atomic level. In MD, Newton's equation of motion is numerically solved to obtain trajectories. These trajectories are used to analyze structural changes, kinetics, and thermodynamic properties of the system. MD is often used to study biological processes such as protein folding/unfolding, ion transport, and enzymatic reactions.

In MD, the position of particle i at time $t + \delta t$, $\vec{r}_i(t + \delta t)$, is calculated using the previously known positions $\vec{r}_i(t)$ at time t and $\vec{r}_i(t - \delta t)$ at time $t - \delta t$ using Verlet equation:[95]

$$\vec{r}_i(t + \delta t) \cong 2\vec{r}_i(t) - \vec{r}_i(t - \delta t) + \delta t^2 \frac{\vec{F}_i(t)}{m_i} \tag{2-36}$$

where m_i is mass of particle and $\vec{F}_i(t)$ is the force acting on it. The force in eq. (2-36) is calculated as the negative gradient of the potential energy, $U(\vec{r})$.

$$\vec{F}_i(t) = -\vec{\nabla}_i U(\vec{r}) \tag{2-37}$$

Depending on how potential energy is calculated, MD simulations are divided into three main categories. They are molecular mechanics MD, quantum mechanics MD, and hybrid quantum mechanics/molecular mechanics MD. Three types of thermodynamic ensembles are commonly simulated.

First, microcanonical (*NVE*), which keeps constant number of particles (*N*), constant volume (*V*), and constant energy (*E*) throughout the MD simulation. This technique is used to calculate dynamically sensitive properties (e.g. diffusion coefficients, dielectric constants, vibrational frequencies etc.). During *NVE* simulations, the system is not coupled to any external variables (thermostat, barostat). The movement of the atoms comes directly from equations (2-36) and (2-37). Second, canonical (*NVT*), which keeps constant number of particles (*N*), constant volume of the system (*V*), and constant temperature of the system (*T*) during the simulations. This ensemble is used preferably to calculate static properties (e.g. radial distribution functions, dipole moment distributions, free energy, hydrogen bonding, heat of vaporization, etc.). In *NVT* simulations, the equations of motion are modified to include coupling to a heat bath, adjusted to keep the temperature of the system around a target value. Third, *NPT*, which has

similar setup to *NVT* except volume of the system, which is allowed to change while keeping constant pressure (*P*). In *NPT* setup, a thermostat is used to keep the temperature around the target value and a barostat is used to keep the pressure around the target value.

2.6 Calculation of infrared spectra

Studying infrared (IR) spectra is very important to understand the microscopic dynamics of the system. As IR spectra are very sensitive to the local environment, the spectral lines give structural information about the system. The computed spectra from MD simulations can also be directly compared with the experimental spectra to validate the accuracy of the potential energy model $U(\vec{r})$. Infrared absorption spectra can be calculated from the time correlation function of the quantum mechanical total dipole moment of the system.[96] In the Schrödinger picture, the infrared absorption coefficient per unit path length, in the limit of weak external field, is computed as[97],

$$\alpha(\omega) = \left[\frac{4\pi^2\omega}{3V\hbar cn(\omega)}\right]\left(1 - e^{-\frac{\hbar\omega}{k_B T}}\right) \times \sum_{f,i} \rho_i \left|\langle\psi_f|\widehat{\boldsymbol{M}}|\psi_i\rangle\right|^2 \delta(\omega_{fi} - \omega) \qquad (2\text{-}38)$$

where $n(\omega)$ is the refractive index as a function of wavelength ω, V is the total volume of the system, $\hbar$ is the reduced Plank's constant, c is the speed of light, and ρ_i is the probability of the system in i^{th} state. $\widehat{\boldsymbol{M}}$ is the total dipole moment operator of the system. ψ_i and ψ_f are initial and final wavefunctions.

With the Heisenberg picture of spectroscopy,[98] eq. (2-38) can be written in terms of the Fourier transform of total dipole moment autocorrelation function:[97]

$$\alpha(\omega) = \left[\frac{4\pi^2\omega}{3V\hbar cn(\omega)}\right]\left(1 - e^{-\frac{\hbar\omega}{k_B T}}\right) \times \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} \langle \widehat{\boldsymbol{M}}(0) \cdot \widehat{\boldsymbol{M}}(t)\rangle dt \qquad (2\text{-}39)$$

$$\alpha(\omega) = \left[\frac{4\pi^2\omega}{3V\hbar cn(\omega)}\right]\left(1 - e^{-\frac{\hbar\omega}{k_B T}}\right) \times I(\omega) \qquad (2\text{-}40)$$

where

$$I(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} \langle \widehat{\boldsymbol{M}}(0) \cdot \widehat{\boldsymbol{M}}(t)\rangle dt \qquad (2\text{-}41)$$

The dipole moment autocorrelation function in eq. (2-41) is entirely calculated from quantum dipole moment operator $\widehat{\boldsymbol{M}}$. However, in practice, the dipole moment autocorrelation function is calculated from classical dipole moment vector $\boldsymbol{M}$. In the classical limit, $I(\omega)$ from eq. (2-41) becomes $I_{\mathrm{cl}}(\omega)$.

$$I_{\mathrm{cl}}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}\langle \boldsymbol{M}(0)\cdot\boldsymbol{M}(t)\rangle dt \tag{2-42}$$

However, to satisfy the "detailed balance" condition

$$I(\omega) = e^{\frac{\hbar\omega}{k_{\mathrm{B}}T}} I(-\omega) \tag{2-43}$$

a quantum correction factor $Q_{\mathrm{QC}}(\omega)$ has to be introduced. The absorption coefficient can then be written as

$$\alpha(\omega) = \left[\frac{4\pi^2\omega}{3V\hbar cn(\omega)}\right]\left(1 - e^{-\frac{\hbar\omega}{k_{\mathrm{B}}T}}\right) \times Q_{\mathrm{QC}}(\omega) \times \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}\langle \boldsymbol{M}(0)\cdot\boldsymbol{M}(t)\rangle dt \tag{2-44}$$

A number of approximations have been proposed for $Q_{\mathrm{QC}}(\omega)$, such as standard approximation (SA), harmonic approximation (HA), Schofield correction (SC), and Egelstaff approximation (EA).[97] Among them, the harmonic correction factor is widely used.[97] HA has the following form:

$$Q_{\mathrm{HA}}(\omega) = \frac{\frac{\hbar\omega}{k_{\mathrm{B}}T}}{1 - e^{-\frac{\hbar\omega}{k_{\mathrm{B}}T}}} \tag{2-45}$$

With eq. (2-45), eq. (2-44) reduces in the following standard form[97,99]

$$\alpha(\omega) = \left[\frac{2\pi\omega^2}{3cVk_{\mathrm{B}}Tn(\omega)}\right] \times \int_{-\infty}^{\infty} e^{-i\omega t}\langle \boldsymbol{M}(0)\cdot\boldsymbol{M}(t)\rangle dt \tag{2-46}$$

In this work, eq. (2-46) is used to compute absorption coefficient based on the autocorrelation function of the total dipole moment of the system. The total dipole moment, volume of the simulation box, and temperature are obtained from the MD simulations.

2.7 Calculation of pK_a

An acid dissociation constant (K_a) measures the strength of an acid in solution. It is also called acidity constant. In general, the acid dissociation reaction can be written as:

$$AH \rightleftharpoons A^- + H^+ \tag{2-47}$$

and the acid dissociation constant (K_a) is defined as the ratio of the product of concentration of conjugate base [A^-] and proton [H^+] to the concentration of compound [AH], i.e.

$$K_a = \frac{[A^-][H^+]}{[AH]} \tag{2-48}$$

For practical reasons, K_a values are often expressed in a logarithmic form. pK_a is defined as the negative logarithm with base 10 of the acid dissociation constant (K_a):

$$pK_a = -\log_{10} K_a \tag{2-49}$$

As K_a, pK_a also characterizes the acidity of the compound. A lower value of the pK_a means that compound is a stronger acid.

pK_a can be calculated using a thermodynamic integration method in which a mechanical constraint is used to fix an order parameter q, defined here as r_{AH}, the distance between the leaving proton and the basic anion (the proton acceptor). For each value of order parameter, the average of Lagrange multiplier λ is calculated from which the mean force $f(q)$ is computed as

$$f(q) = \langle \lambda \rangle_q - \frac{2k_B T}{q} \tag{2-50}$$

The potential of mean force $w(q)$ is then computed relative to its value $w(q_0)$ at some reference value q_0 as

$$w(q) = -\int_{q_0}^{q} f(q')dq' \tag{2-51}$$

For $q = r_{AH}$, the inverse of the dissociation constant as a function of the integration parameter R_C for the reaction AH ⇌ A^- + H^+ is given by[100,101]

$$K^{-1}(R_c) = C_0 \int_0^{R_c} 4\pi r_{AH}^2 \, e^{-\frac{w(r_{AH})}{k_B T}} dr_{AH} \tag{2-52}$$

where C_0 is the standard concentration equal to 1M. The dissociation constant $K_a(R_c)$ of an acid in water is computed as[100,101]

$$K_{\mathrm{a}}(R_{\mathrm{c}}) = \frac{\alpha(R_{\mathrm{c}})^2}{1-\alpha(R_{\mathrm{c}})}\,\frac{N}{C_0 V} \tag{2-53}$$

where N the number of reactive sites and $\alpha(R_{\mathrm{c}})$ is the dissociation fraction given by[100,101]

$$\alpha(R_{\mathrm{c}}) = 1 - \frac{\int_0^{R_c} 4\pi r_{\mathrm{AH}}^2\, e^{-\frac{w(r_{\mathrm{AH}})}{k_{\mathrm{B}}T}}\, dr_{\mathrm{AH}}}{\int_0^{(V^{1/3})/2} 4\pi r_{\mathrm{AH}}^2\, e^{-\frac{w(r_{\mathrm{AH}})}{k_{\mathrm{B}}T}}\, dr_{\mathrm{AH}}} \tag{2-54}$$

For clean water with N water molecules, the dissociation constant $K_{\mathrm{w}}(R_{\mathrm{c}})$ is computed as[100,101]

$$K_{\mathrm{w}}(R_{\mathrm{c}}) = \alpha(R_{\mathrm{c}})\,\frac{N}{C_0 V} \tag{2-55}$$

Finally, the following formula is used to estimate the pK_{a} values.

$$\begin{aligned} \mathrm{p}K_{\mathrm{a}} &= -\log(K_{\mathrm{a}}) \\ \mathrm{p}K_{\mathrm{w}} &= -\log(K_{\mathrm{w}}) \end{aligned} \tag{2-56}$$

The values of pK_{a} and pK_{w}, depend on the choice of R_{c}. The value of R_{c} can be established using the pK_{a} value of a known acid, and used to estimate the pK_{a} of a substituted but similar acid, or the pK_{a} of the same acid in a different chemical environment. This procedure allows us to calculate, for instance, pK_{a} shifts induced by metal ligation.

3 Re-parameterization of AM1 semiempirical model for liquid water: hydrogen bonding and proton transfer reactions

3.1 Abstract

Because of their computational cost, applications of high-level quantum mechanical methods are limited to small or medium sized molecular systems. Semiempirical (SE) methods based on the neglect of diatomic differential overlap approximation represent an acceptable computational cost even for relatively large systems, provided they are tailored to reproduce the physical and chemical properties of interest. In this work, the standard semiempirical AM1 model is re-parameterized for describing the properties of liquid water and the structure and energetics of proton transfer reactions in water. SE parameters for H and O are optimized using a genetic algorithm to reproduce the geometries and energies of a set of compounds calculated from density functional theory, with a special emphasis on the hydrogen bonding geometries. The training set contains all conformations and transition structures of the water monomer, dimer, and trimer in their neutral, protonated, and deprotonated forms. The transferability of the optimized model is assessed for a large testing set of structures of water tetramers, pentamers, and hexamers. The optimized model, called AM1-LW, provides significant improvement over existing standard SE models. In contrast to other SE models, the AM1-LW model yields liquid water properties consistent with experimental data, including radial distribution functions, enthalpy of vaporization, self-diffusion coefficient, dielectric constant, and Debye relaxation time. Unlike AM1, AM1-W, and AM1PG-W models, the AM1-LW model also reproduces the infrared spectrum of liquid water with good agreement to experiment. The gas-phase and liquid-phase results from AM1-LW suggest that the model offers a good alternative to study proton transfer reactions in water without altering the theoretical framework of standard SE models.

3.2 Introduction

Water is (arguably) the most important protic solvent in physical, chemical, and biological systems.[2] It plays a key role in acid-base reactions in solution and participates in proton transfer reactions in biological systems by connecting proton donors and acceptors through hydrogen-bonded networks and by changing energy level of solvated species.[2,16,102,103] It remains a major

challenge to develop models of water that describe its unusual thermodynamics and dynamics as a neat liquid,[5] as well as its protic character and its properties as a solvent for H^+ and OH^- ions.

Proton transfer mechanisms in water have been extensively studied,[13,23–25] however, water-mediated proton transfer reactions have been investigated only for a relatively small number of systems[16,104–106]. Since the mechanism of water-mediated proton transfer reactions depends on the specific arrangement of the acid and the base (and the hydrogen-bonded network of water molecules between the two), it is essential to develop a more practical modeling approach that is both computationally efficient and reliable for the purpose of describing water-mediated proton transfer reactions in various environments.

High-level ab initio molecular dynamic (MD) simulations often yield some of the structural and dynamical properties and electronic structure of the liquid water in line with experimental values.[107–111] However, due to the high computational cost of ab initio MD simulations, their application is limited to small to moderate sized systems.[112] Despite the popularity of force fields[113–115], they are inapplicable on electronic transition, electronic transport, and the reaction systems with bond breaking or forming.

Semiempirical (SE) methods retain the salient features of quantum chemistry as they provide a description of the electronic structure and allow covalent bonds to be broken or formed while being computationally efficient even for large-scale systems. SE methods use quantum mechanical formulation to define the potential energy functions. They approximate many of two-electron integrals by introducing external empirical parameters and thus expedite the numerical computations. These parameters are obtained by fitting empirical data from experimental and/or high-level ab initio methods. Because SE methods are computationally efficient, they can be used in molecular modeling of reactions, industrial designing of chemical syntheses, developing and evaluating of different methodologies and algorithms, gaining insight about electronic properties of any complexes, etc.[116] SE methods can also be employed in quantum mechanical molecular mechanics (QM/MM) setup for examining various properties of very large systems of chemical or biological phenomena.[117,118]

Standard SE methods such as AM1,[73] PM3,[86] and PM6[119] are based on neglect of diatomic differential overlap (NDDO)[72] approximations. These methods were widely used in the past, are

currently being used and improved. Re-parameterization of the SE molecular orbital methods is often required to overcome the shortcomings resulting from the inherent approximations of the methods. When properly parameterized, some of these methods have been shown to accurately describe molecular structure and non-bonded interactions.[120–124] New SE methods are currently being developed and applied to study structures and thermodynamics of large complexes including proteins.[93,125–129]

Traditionally, SE models have been parameterized to describe a few properties for a wide variety of organic compounds. Early SE models including the RM1 method[130] were parameterized using this technique. For instance, RM1 has been parameterized to describe heat of formation, dipole moments, ionization potentials, and geometric variables (bonds and angles) of a large training set of 1736 organic and bioorganic complexes. More recently, new SE models are being developed and/or improved the existing SE models to describe properties for a specific class of systems.[93,123,124,127] Since the original purpose of the parameterization of SE models is to reproduce the structural and energetic properties of small organic clusters, they fail to reproduce the structures and interaction energies of bioinorganic clusters. For example, AM1 and PM3 work well for describing Zn–N bonds of model complexes for carbonic anhydrase but does not reproduce the coordination number of sterically crowded Zn–O complexes and underestimate Zn–O interaction energies.[131] Recent study shows that the re-parameterized SE methods (e.g. $RM1_{BH}$) can be used to investigate the weak interaction of biological systems especially binding energy calculation of hydrogen bonds.[88] With an emperical dispersion correction, SE models (AM1-D and PM3-D) improve the interaction energies of hydrogen-bonded model complexes for DNA pairs.[132] The system-specific semiempirical methods are being currently developed to describe biological interactions accurately.[88,118,121,132,133]

Most of the studies showed that the AM1 method does not correctly represent hydrogen-bonding for hydrogen bonded systems.[134–137] Water dimer, in ground state, has trans linear hydrogen-bonded configuration that is verified both experimentally[138,139] and theoretically[140,141]. AM1, however gives a bifurcated structure in which the two hydrogen atoms of one water molecule coordinate the lone pair of the other one.[142–144] As a result, molecular dynamic simulations of AM1 do not yield liquid water structure, coordination number, enthalpy of vaporization and condensed-phase dipole moment consistent with experimental values.[145] Like

AM1, PM3 and PM6 do not reproduce condensed-phase water properties either.[146] Several attempts have been done to improve the performance of SE methods for liquid water but only a small number of models capture only few properties of water.[118,124,145–147] The performance of SE models on hydrogen-bonded systems can be improved by parameterizing new core-core repulsion function.[89,90,145]

Since SE models are highly simplified quantum mechanical descriptions, their transferability to complexes and reactions falling outside the scope of the original parameterization is unreliable at best. For such systems, reaction-specific SE models need to be developed. Reaction-specific SE models usually address some of the poor transferability of standard SE methods.[123] Some of the models describe the gas-phase proton transfer profiles but fail to retain the correct structural description of bulk water.[124,148] Wang *et al.* have reported an approach for semiempirical parameterization based on genetic algorithms.[120] This technique was used to develop two models for proton transfer reactions in water clusters: AM1-W, a re-parameterization of the original AM1 model for water clusters, and AM1PG-W, a parameterization of an AM1-like model with a core repulsion function containing additional pairwise Gaussians.[120] Both AM1-W and AM1PG-W models capture hydrogen-bonding energies and energy profiles of proton transfer reactions in gas phase water clusters better than AM1, PM3, and RM1 models. However, these models are not designed for capturing the properties of bulk liquid water. Therefore, a new SE model has to be developed that reproduce both gas phase properties and liquid dynamics of water.

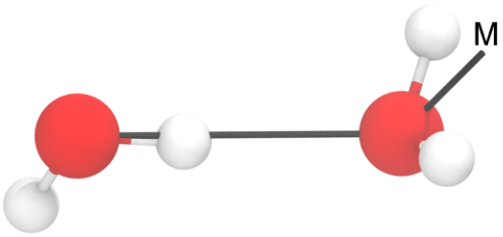
In this chapter, we report the re-parameterization of the original AM1 for describing gas-phase proton transfer reactions and liquid water properties using the parameterization technique reported in our previous work.[120] We use the same training and testing sets as in ref. [120] but we modify the error function to give higher importance to hydrogen-bonding geometry and yield improved liquid water structure. The transferability of the model is assessed for larger cluster systems that mimic biological proton transfer and/or proton transport processes. The model is further employed in molecular dynamic simulations to measure its capability at reproducing static (radial distribution functions, dipole moment distributions, enthalpies of vaporization, and hydrogen bonds) and dynamic (self-diffusion coefficients, dielectric constants, autocorrelation functions, and IR spectra) properties of liquid water.

3.3 Computational Methods

3.3.1 Error function and parameterization procedure

The semiempirical model using AM1 Hamiltonian[73] was re-parameterized to reproduce the structural and energetic properties of water clusters as well as gas-phase proton transfer reactions. Special attention was paid to the geometry of the water dimer, which is critical for obtaining correct hydrogen-bonding structure in the liquid phase. AM1 parameters for hydrogen and oxygen were derived by fitting a set of properties obtained from hybrid DFT calculations at the B3LYP/6-311++G(2d,2p) level, which has been shown to give acceptable energetics for water clusters[149] and is much faster than high-level methods such as MP2 and CCSDT. All DFT calculations were performed using Gaussian 09[150] and semiempirical calculations were done using an in-house version of MOPAC 07.[151] Parameterization was carried out using a set of small water clusters (here onwards *training set*), which consists of all possible stable conformations of water monomer, dimer and trimer in their neutral, protonated, and deprotonated forms. Geometries, proton transfer energies, proton affinities, transition state reactions, hydrogen bonding energies, and reaction energies were considered.

Liquid water structure is highly sensitive to the hydrogen-bonding geometry, but AM1-W and AM1PG-W models were parameterized without any special consideration for these properties. As shown in Figure 3.1, Both AM1-W and AM1PG-W underestimate oxygen-oxygen distance of water dimer compared to DFT value. For ∠OOM (angle between O···O and bisector (O–M) of angle H–O–H of the H-bond receptor water), AM1-W overestimates and AM1PG-W underestimates. To get a reasonable liquid structure, explicit dimer bond and angle contributions were included in the error function.



O-O	∠OOM	
2.72 Å	128.6°	AM1-W
2.65 Å	115.9°	AM1PG-W
2.92 Å	121.2°	B3LYP
2.77 Å	121.8°	AM1-LW (this work)

Figure 3.1: Oxygen-oxygen bond distance and ∠OOM (where M is a point on the HOH bisector of the hydrogen-bond accepting water) of water dimer obtained from AM1-W, AM1PG-W, B3LYP, and the AM1-LW models.

The error function χ is defined as the sum of deviations of various structural and energetic properties from their reference values. The properties include overall cluster structures (for mean square deviations (*MSD*)), bond distances (*B*) and angle (*A*) in water dimer, reaction energies (E_R), proton affinities (E_{PA}), activation energies (E_A), hydrogen bonding energies (E_{HB}), and proton transfer energies (E_{PT}). The mathematical form of χ is given by:

$$\begin{aligned}\chi = {} & W_{MSD}\sum_i |MSD_i| + W_B\sum_i \left|B_{dimer}^{model}(i) - B_{dimer}^{ref}(i)\right|^2 \\ & + W_A\left|A_{dimer}^{model}(i) - A_{dimer}^{ref}(i)\right|^2 + \sum_i \left|E_R^{model}(i) - E_R^{ref}(i)\right| \\ & + \sum_i \left|E_{PA}^{model}(i) - E_{PA}^{ref}(i)\right| + \sum_i \left|E_A^{model}(i) - E_A^{ref}(i)\right| \\ & + \sum_i \left|E_{HB}^{model}(i) - E_{HB}^{ref}(i)\right| + W_{PT}\sum_i \left[\frac{1}{n_{PT}}\sum_{k=1}^{n_{PT}} \left|E_{PT}^{model}(i,k) - E_{PT}^{ref}(i,k)\right|\right]\end{aligned} \quad (3\text{-}1)$$

where W_{MSD} (= 33 kcal/mol.Å^2), W_B (= 250 kcal/mol.Å^2), W_A (= 500 kcal/mol.rad^2), and W_{PT} (= 2) are weight factors for corresponding quantities. Superscripts "model" and "ref" represent the calculated respective properties using optimized SE model and reference DFT results. n_{PT} (= 5) is the number of points in each proton transfer (PT) profile. Proton transfer energies were calculated by scanning the H(donor)···O(acceptor) distance from 1.7 to 0.9 Å in steps of –0.2 Å while keeping O(donor), H(donor), and O(acceptor) collinear. $E(i,k)$ represents the energy of structure i at position k (=1 to 5) along the proton transfer energy scan.

Mean square deviations were calculated by aligning calculated structure from SE model to reference structure from DFT for all non-equivalent structures except OH^-. In the error function, the covalent O–H (donor) and noncovalent H(donor)···O(acceptor) bond lengths and the O (donor)–H(donor)···O(acceptor) angle of water dimer ($HOH\cdots OH_2$) were included. Large weighting factors for bonds (W_B) and angle (W_A) were used to emphasize dimer structure. Reaction energies, proton affinities, activation energies, hydrogen bonding energies and proton transfer energies are calculated as previously.[120] In MOPAC, geometry optimization termination criterion is set to GNORM = 1.0 kcal/mol/Å for the parametrization and to GNORM = 0.01 kcal/mol/Å for the final assessment for all models (and for vibrational frequency calculations).

A genetic algorithm approach was used to optimize the SE parameters set by using a parallel version of the PIKAIA program[94]. In a single PIKAIA run, the evolution of 100 number of individuals (each individual represents a set of SE parameters) in a population (within the search regions) were used for 300 numbers of generations. For each set of parameters, the fitness score is evaluated for the natural selection by fitness function, $1/(\chi + 1)$. The set of parameters with highest score is collected in each PIKAIA run. A new generation of individuals is generated by breeding two good "parent" individuals. To increase the genetic diversity, new generations were obtained by the genetic crossover of the previous pair of individuals, followed by random mutation. 18 oxygen parameters were optimized followed by 14 hydrogen parameters. Parameters were initially set to their original AM1 values[73](also see Table 3.1) and were allowed to change up to ±50% from their initial values. Three rounds of optimization, for each element, were needed to minimize the error value. The error value is further minimized by narrowing-down the search region, allowing only ±20% change from the previous-round values. Three more rounds were performed in a similar way, over progressively narrower regions allowing ±10%, ±5%, and ±1% change from previous values. Two rounds of optimization were needed to converge the error value. Finally, no improvement on the error value was observed and the parameterization was terminated. The transferability of the model was assessed for a large set of water clusters (here onwards *testing set*), which includes large water clusters of tetramers, pentamers, and hexamers in their neutral, protonated, and deprotonated forms. Exactly same training and testing sets are used for the parameterization as described in our previous work.[120]

3.3.2 Molecular dynamics simulations

Semiempirical molecular dynamics (MD) simulations of liquid water were performed using the CP2K package[152] for four different models namely original AM1, AM1-W, AM1PG-W and the optimized model. The CP2K code was modified for AM1PG-W implementation. For each SE model, 100-ps long MD simulations were performed in the canonical ensemble (*NVT*) with a time step of 0.5 fs. The first 10-ps of each simulation were discarded and the last 90-ps were analyzed.

Static properties such as radial distribution functions, dipole moment distributions, enthalpy of vaporization, and hydrogen bonds were calculated from canonical ensemble (*NVT*) whereas dynamic sensitive properties namely self-diffusion coefficients, dielectric constants, dipole moment-time correlations, and IR spectra were calculated from microcanonical (*NVE*) ensemble. All MD simulations were carried out for 64 water molecules in a periodic cubic box of L = 12.4138 Å, corresponding to experimental density ρ = 1.0 g/cm^3. The wave function convergence criterion was set to 10^{-7} au. Periodic Ewald summation was used to treat the long-range electrostatic interaction with α parameter of 0.5 Å^{-1} and number of grid points of 25 for all directions with quadrupole expansion. The standard NDDO Klopman-Dewar-Sabelli-Ohno (KDSO)[153–157] equation was used for the screening of the Coulomb interactions with interactions up to half of the box (L/2 = 6.2069 Å). In *NVT* simulations, massive Nosé-Hoover thermostat with a chain length of 3 and a time constant of 2000 fs were applied to all degrees of freedom to maintain constant temperature of 300 K. Same cutoff schemes as in ref. [146] were used for the MD simulations.

10 independent multiple *NVE* simulations of 40 ps each (400 ps in total) were performed using each SE model (as explained above) from different starting points. The initial configurations for each 10 *NVE* simulations were selected from 10th, 20th, …,90th, and 100th ps points of the 100-ps parent *NVT* simulations and dynamical properties were calculated and averaged. These multiple independent simulations were used to measure the convergence of the different properties.

Table 3.1: Optimized AM1-LW parameters compared with original AM1[73] and AM1-W[120] models. Symbols have their usual meaning, as defined in references [73] and [120].

	AM1		AM1-W		AM1-LW (this work)	
Parameter	**H**	**O**	**H**	**O**	**H**	**O**
U_{ss} (eV)	−11.396427	−97.83000	−12.1942	−128.1573	−12.560002	−124.312581
U_{pp} (eV)		−78.26238		−79.8276		−79.588146
ζ_s (au)	1.188078	3.108032	1.0574	3.4188	1.078537	3.497468
ζ_p (au)		2.524039		2.5493		2.528886
β_s (eV)	−6.173787	−29.27277	−5.0625	−36.2982	−4.860006	−37.859064
β_p (eV)		−29.27277		−38.9328		−37.180813
G_{ss} (eV)	12.848	15.42	14.7752	21.4338	14.331944	21.583836
G_{sp} (eV)		14.48		18.6792		18.193541
G_{pp} (eV)		14.52		17.5692		17.042124
G_{p2} (eV)		12.98		9.8648		10.338310
H_{sp} (eV)		3.94		2.8368		2.811269
α (Å^{-1})	2.882324	4.455371	2.7670	5.7029	2.739361	5.799824
K_1 (eV)	0.122796	0.280962	0.0847	0.3484	0.049143	0.352574
L_1 (Å^{-1})	5.0000	5.0000	6.6995	3.0000	3.953000	3.030000
M_1 (Å)	1.2000	0.847918	1.0320	1.0005	0.980400	0.974529
K_2 (eV)	0.00509	0.08143	0.0038	0.0497	0.003321	0.048729
L_2 (Å^{-1})	5.0000	7.0000	6.4500	6.0900	8.256000	6.150900
M_2 (Å)	1.8000	1.445071	1.6020	1.9364	1.345680	1.841512
K_3 (eV)	−0.018336		−0.0260		−0.021350	
L_3 (Å^{-1})	2.0000		2.0400		1.897200	
M_3 (Å)	2.1000		1.1760		1.270080	

3.4 Results and discussion

The optimized parameters for hydrogen and oxygen are presented in Table 3.1. The optimized SE model is called AM1-LW ("AM1 for Liquid Water"). Original AM1 and AM1-W parameters are also presented for comparison. The AM1-LW parameters deviate from the standard AM1 values by 25% on average for hydrogen parameters and 23% on average for oxygen parameters, and by more than 50% for only two parameters (K_1 and L_2 for hydrogen). Large deviations have been reported in the literature for similar re-parameterization attempts. For example, in a force-matching re-parameterization of the PM6 model, Welborn *et al.*[147] found that the α parameter of hydrogen deviated by 631% from its original PM6 value[119].

3.4.1 Performance on gas-phase clusters

As expected, the AM1-LW model performs better than other models for all properties of the training set. The AM1-LW model slightly underestimates an oxygen-oxygen distance (2.77 Å) of water dimer but it gives an O···O–M angle (see Figure 3.1) of 121.8° consistent with the B3LYP result of 121.2°. The performance of the AM1-LW model is compared with standard AM1, PM3 and RM1 models using DFT results as reference. Figure 3.2 reports the various energies associated with proton transfer reactions, the main target of this work. As shown in Figure 3.2(a) (training set) and Table 3.2, none of the AM1, PM3 and RM1 models satisfactorily reproduce proton affinities. The mean absolute error (MAE) for proton affinities is 1.66 kcal/mol for the AM1-LW model, but 20.78 kcal/mol for AM1, 15.37 kcal/mol for PM3, and 29.74 kcal/mol for RM1. The AM1-LW model also gives small MAEs for both proton transfer energies (1.99 kcal/mol) and activation energies (1.81 kcal/mol).

The transferability of the model is evaluated using the testing set. As shown in Figure 3.2(d-f) and Table 3.2, the AM1-LW model shows significant improvement over AM1, PM3, and RM1. Both AM1 and RM1 have very large MAEs on proton affinity, proton transfer and activation energies. Overall, PM3 performs better on proton transfer and activation energies but fails to reproduce the proton affinities. Unlike other models, the AM1-LW model has excellent correlation with DFT (see Figure 3.2) on proton affinities, proton transfer, and activation energies with an overall MAE of 1.94 kcal/mol. The AM1-LW model underestimates PT energies for 5 of

the 50 points along the 10 proton transfer profiles of the testing set (see Figure 3.2(e)). These correspond to the 0.9-Å point of the PT profiles of protonated tetramers (two isomers) and pentamer and to the 0.9 and 1.1-Å points of the PT profile of protonated hexamer. The large errors in PT energies of these specific systems[158] are associated with weak hydrogen bonding coming from the non-linearity of hydrogen bonds on the local structures. These large errors stem from the clusters adopting incorrect hydrogen-bonding structures once the excess proton is moved too close to its acceptor molecule. When proton is closed to the acceptor water, AM1-LW fails to retain linearity of hydrogen bonds between proton donor and acceptor water molecules during proton transfer scan. In case of protonated hexamer, it also loses its initial Eigen conformation to adopt a conformation in which the two water molecules not directly involved in the proton transfer form a hydrogen bond. The large errors in energies coming from the breaking old and forming new hydrogen bonding were also previously reported.[120] only AM1PG-W model reproduces these specific PT profile.[120] The gas-phase performance of AM1-W and AM1PG-W models are explained in details in our previous work.[120]

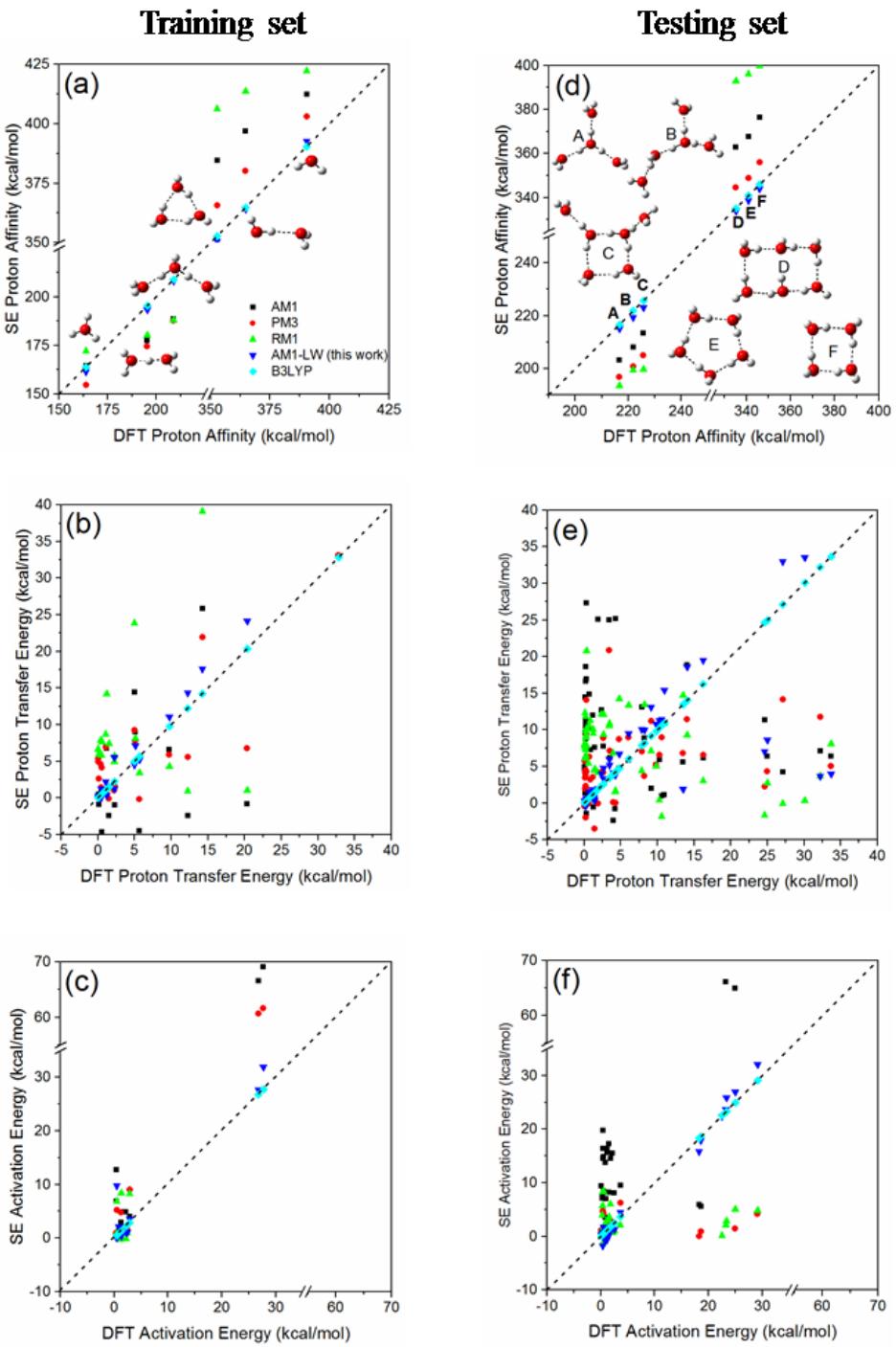


Figure 3.2: Performance of AM1, PM3, RM1, and AM1-LW (this work) on training set (panels a to c) and testing set (panels d to f). SE results are plotted against DFT results and diagonal dotted lines indicate perfect correlation. The cluster structures for which proton affinities were calculated are shown inside panels a and d.

Table 3.2: Mean absolute errors from AM1, PM3, RM1, and AM1-LW (this work) relative to DFT results. *N* is the number of corresponding terms used in the error function. Combined average error (average of average training and testing sets errors) is also reported.

			Mean absolute errors (kcal/mol)			
		N	AM1	PM3	RM1	AM1-LW (this work)
Training set	Proton affinity	6	20.78	15.37	29.74	1.66
	Proton transfer energy	4×5=20	6.05	4.13	11.26	1.99
	Activation energy	12	9.27	7.15	19.67	1.81
Testing set	Proton affinity	6	20.68	14.86	39.68	1.98
	Proton transfer energy	10×5=50	12.86	6.75	9.61	3.20
	Activation energy	28	19.42	9.16	13.09	0.98
Combined average error			**14.84**	**9.57**	**20.51**	**1.94**

3.4.2 Structure of liquid water

Figure 3.3 presents radial distribution functions (RDFs) of liquid water simulated with AM1, AM1-W, AM1PG-W, and AM1-LW models in comparison with neutron diffraction experiment.[159] The locations and amplitudes of different features of the RDFs, as well as the coordination numbers (CNs), are presented in Table 3.3. CNs are calculated by integrating $g_{\mathrm{OO}}(r)$ from origin to first minimum. As shown in Figure 3.3(a) and Table 3.3, the AM1, AM1-W, and AM1PG-W models do not reproduce experimental $g_{\mathrm{OO}}(r)$ profile correctly. AM1 gives a first shell centered at 2.77 Å in line with experiment but fails to reproduce the positions of the second and third shells. Moreover, the first peak of $g_{\mathrm{OO}}(r)$ is too broad and extends unusually far ($r_{\mathrm{OO}}^{\mathrm{min1}}$ = 4.06 Å), which results in a high O–O coordination number of 10.40, compared to the experimental value of 4.5.[110,159] AM1-W gives sharp "ice-like" $g_{\mathrm{OO}}(r)$, yielding a low O-O

coordination number of 3.72, the overall signature of a very rigid hydrogen bond network. AM1PG-W gives a split first shell coming from a mixture of linear hydrogen bonds (with r_{OO} ~ 2.6 Å and r_{OH} ~ 1.7 Å) and of bifurcated hydrogen bonds in which the two molecules have their dipoles aligned (with r_{OO} ~ 3.1 Å and r_{OH} ~ 2.3 Å). The broadening of first shell results in a large coordination number of 11.04 from AM1PG-W. By contrast, AM1-LW gives $g_{\mathrm{OO}}(r)$ consistent with experiment with a first peak, $r_{\mathrm{OO}}^{\mathrm{max1}}$ around 2.67 Å and $g_{\mathrm{OO}}^{\mathrm{max1}}$ of 2.78 in good agreement with experimental position of 2.73 Å and $g_{\mathrm{OO}}(r)$ of 2.75 (see Figure 3.3(a) and Table 3.3). The model also returns both $r_{\mathrm{OO}}^{\mathrm{max2}}$ and $g_{\mathrm{OO}}^{\mathrm{max2}}$ values accurately in comparison with experimental values for second shell as well. The estimated first-shell O–O coordination number from the AM1-LW model is 4.85 in good agreement with experimental value of 4.5.[110,159]

Distributions $g_{\mathrm{OH}}(r)$ and $g_{\mathrm{HH}}(r)$ for AM1 (Figure 3.3(b) and Figure 3.3(c)) are consistent with a liquid that forms weak hydrogen bonds, with a strongly shifted and poorly-defined first peak for $g_{\mathrm{OH}}(r)$ and a broad $g_{\mathrm{HH}}(r)$ first peak. AM1-W gives strongly ordered hydrogen bonding characteristics. AM1-LW shows balanced hydrogen bonded structure of water in a good agreement with experiment.

Because many properties of water stem from the structure and stability of its hydrogen bond network, it is important to check which models yield a correct number of hydrogen bonds. Two water molecules are assumed to be hydrogen bonded when the distance between two oxygen atoms is less than or equal to 3.5 Å and the O–H···O angle is less than or equal to 45°.[160] As shown in Table 3.3, AM1 and AM1PG-W give small number of hydrogen bonds per water molecule. AM1-W (3.82) and AM1-LW (3.30) yield average hydrogen bonds per water molecules in a good agreement with experimental value of 3.58[161] than AM1 (2.24) and AM1PG-W (2.44). AM1-LW shows a good improvement in describing general solvation structure of liquid water by reproducing collective features of $g_{\mathrm{OO}}(r)$, $g_{\mathrm{OH}}(r)$, and $g_{\mathrm{HH}}(r)$.

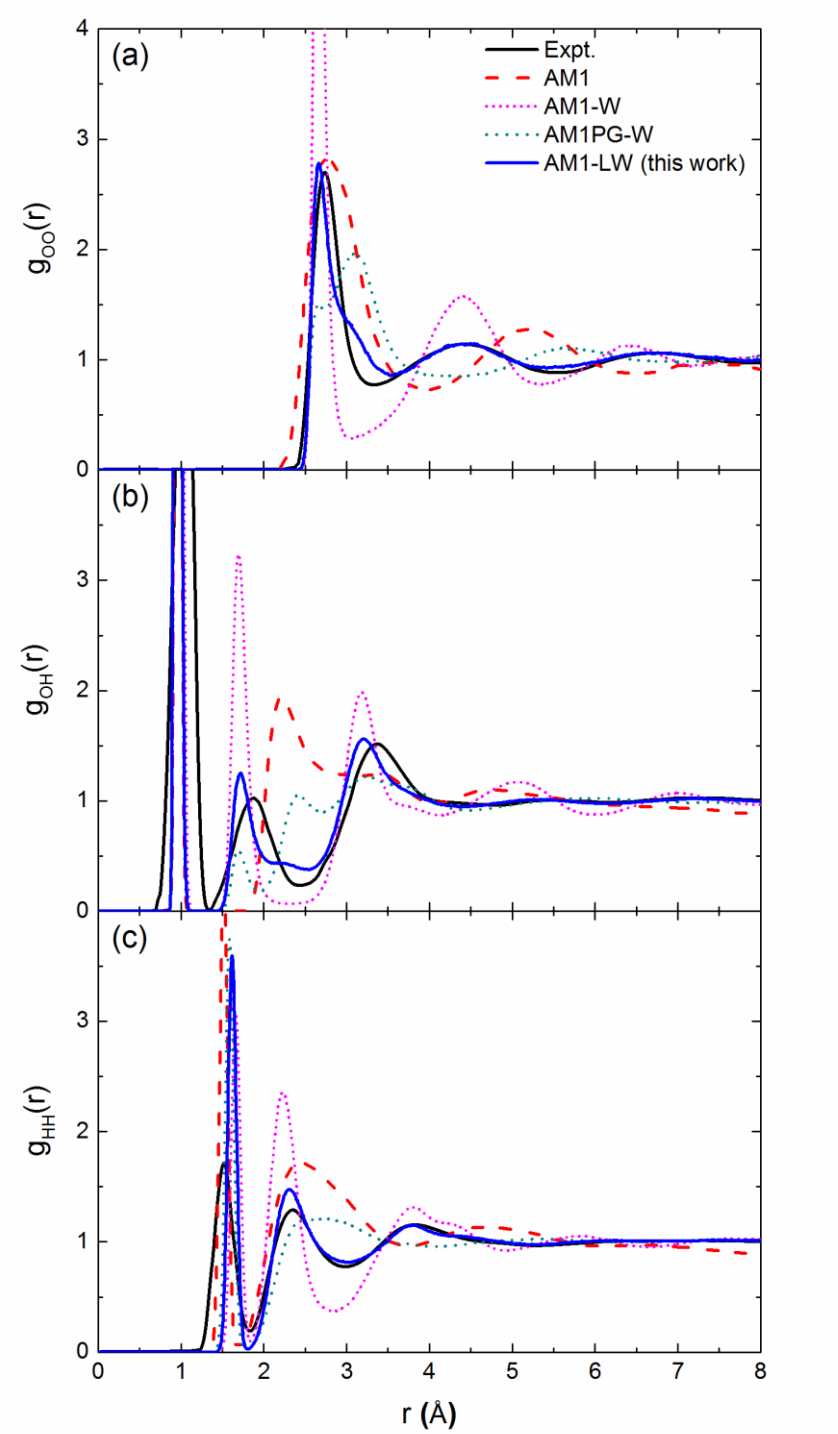


Figure 3.3: Radial distribution functions for OO, OH, and HH pairs of liquid water from AM1, AM1-W, AM1PG-W and AM1-LW (this work) models. Experimental data taken from ref. [159] are also included for the comparison.

Table 3.3: Results of *NVT* simulations of liquid water and comparison to experiment. Experimental RDF data are taken from ref. [159]

Methods	r_{OO}^{max1} [Å]	g_{OO}^{max1}	r_{OO}^{min1} [Å]	g_{OO}^{min1}	r_{OO}^{max2} [Å]	g_{OO}^{max2}	CN	CN_{HB}	ΔH_{vap} [kcal/mol]
Expt.	**2.73**	**2.75**	**3.36**	**0.78**	**4.52**	**1.16**	**4.5**[a]	**3.58**[b]	**10.5 ± 0.3**[c]
AM1	2.77	2.85	4.06	0.73	5.19	1.28	10.40	2.24	9.0
AM1-W	2.66	6.93	3.06	0.29	4.40	1.58	3.72	3.82	17.3
AM1PG-W	3.10	1.97	4.38	0.84	5.74	1.11	11.04	2.44	8.0
AM1-LW (this work)	**2.67**	**2.78**	**3.69**	**0.89**	**4.53**	**1.15**	**4.85**	**3.30**	**12.1**

a. Reference [110,159] b. Reference [161] (at T = 298 K) c. Reference [162] (at T = 25°)

3.4.3 Enthalpy of vaporization

Enthalpy of vaporization is calculated from average intermolecular interaction energy E_{int} using the following formula:[145,163]

$$\Delta H_{\text{vap}} = -E_{\text{int}} + RT \tag{3-2}$$

where R is the gas constant and E_{int} is defined in terms of E_n, the average potential energy of n interacting water molecules during a molecular dynamics simulations of liquid water, and of E_1, the potential energy of single water in gas phase, and is calculated as:

$$E_{\text{int}} = \frac{\langle E_n \rangle - n\,E_1}{n} \tag{3-3}$$

As shown in Table 3.3, enthalpies of vaporization are within ±2.5 kcal/mol of the experimental value of 10.5±0.3 kcal/mol[162] for all models except AM1-W. AM1-W (17.3 kcal/mol) significantly overestimates the enthalpy of vaporization compared to experimental value.

3.4.4 Dipole moment distribution in liquid water

To check the polarization effects in condensed phase, dipole moments of water monomer and dimer in gas-phase and monomer in liquid water are analyzed. The average dipole moment of monomer in liquid water is calculated from *NVT* MD simulations.

Table 3.4 summarizes the calculated dipole moments compared to experiment and Figure 3.4 shows the dipole moment distribution of water monomer in liquid water as predicted from AM1, AM1-W, AM1PG-W, and AM1-LW models. While all models describe some amount of molecular polarization, some models get polarized more. During the development of AM1 model, dipole moment of water was included in the error function. As a result, with no surprise, AM1 yields the gas-phase dipole moment of water monomer but does poorly for the liquid-phase. As shown in Table 3.4, despite reproducing the dipole moment of water monomer, AM1 model underestimates for both gas-phase water dimer (2.32 D) and water monomer (2.04 D) in liquid water. AM1-W model gets polarized the most, going from 2.75 D in gas phase to 3.40 D in liquid phase, likely due to its overly strong hydrogen-bonding character.[164] AM1PG-W model however overestimates dipole moment for gas-phase water monomer, but yields dipole moment of water dimer and monomer in liquid water of values 2.66 and 2.97 D in consistent with experimental value of 2.64 and 2.90 D respectively. It is noticed that the small O···O–M angle in water dimer (see Figure 3.1) from AM1PG-W causes decrease in dipole of water dimer (2.66 D) than water monomer (2.71 D). The AM1-LW slightly overestimates the dipole moment of gas-phase water monomer and dimer and yields dipole moment of monomer in liquid that lie between AM1-W and AM1PG-W values as shown in Figure 3.4.

Table 3.4: Gas-phase dipole moment of water monomer and dimer and dipole moment of monomer in liquid phase.

Methods	Dipole moment (Debye)		
	Monomer	Dimer	Monomer in liquid
Expt.	1.85[a]	2.64[b]	2.9 ± 0.6[c]
AM1	1.86	2.32	2.04
AM1-W	2.75	3.45	3.40
AM1PG-W	2.71	2.66	2.97
AM1-LW (this work)	2.75	3.05	3.18

a) Reference [165]

b) Reference [166]

c) Based on partial atomic charges estimated from X-ray diffraction data [167]

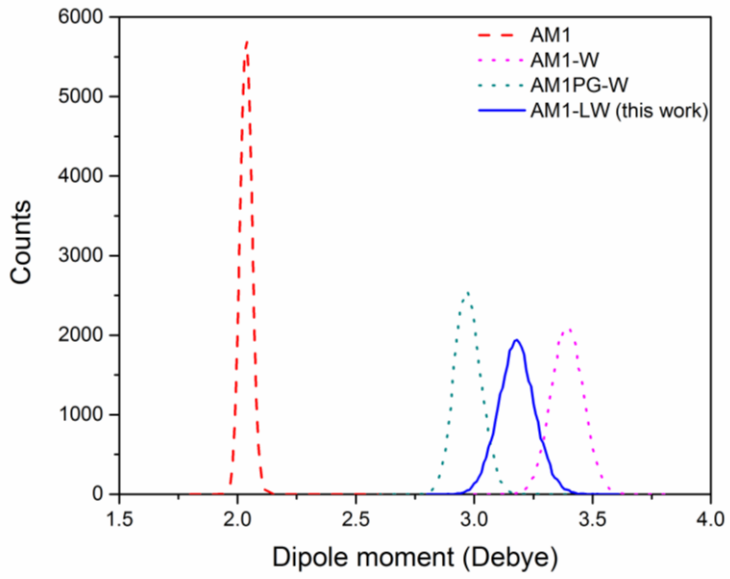


Figure 3.4: Distribution of the monomer dipole moment in liquid water at $T = 300$ K.

3.4.5 Self-diffusion coefficient

Self-diffusion coefficient of liquid water is calculated from the long-time limit of the mean square displacement of the oxygen atoms, using the following formula:

$$D_{\mathrm{PBC}} = \lim_{t\to\infty} \frac{1}{6t} \langle \frac{1}{N} \sum_{i=1}^{N} \left[\boldsymbol{r}_{\mathrm{O},i}(t) - \boldsymbol{r}_{\mathrm{O},i}(0) \right]^2 \rangle \qquad (3\text{-}4)$$

Mean square deviation over the delay time is calculated from the last 30-ps of each 40-ps microcanonical simulation. Since self-diffusion coefficient measures the linear growth of mean square deviation, the t < 5 ps range and the poorly sampled t > 25 ps range are ignored for the linear fit. The self-diffusion coefficients D_{PBC} obtained from AM1, AM1-W, AM1PG-W, and AM1-LW (this work) are presented in Table 3.5. As diffusivity of the liquid strongly depends on the simulation system size, it is requisite to incorporate finite-size effects on D_{PBC}.[168] Since we do not have access to the viscosities of the different models (which are needed to calculate finite system-size correction factors[168]), it is more straightforward to compare our D_{PBC} results with the D_{PBC} value to be expected from the experimental diffusion coefficient and viscosity. As shown in Table 3.5, the self-diffusion coefficient of water from AM1-LW (0.37 Å^2/ps) is higher than the experimental value (0.18 Å^2/ps). AM1 and AM1PG-W give high self-diffusion coefficient. The low value of self-diffusion coefficient from the AM1-W reflects the condensed-phase of water is overly structured. The sharp RDFs in Figure 3.3 also support this conclusion. AM1-LW, however, shows great improvement over other SE models, including PM3 (0.78 Å^2/ps) and PM6 (1.28 Å^2/ps).[146]

Table 3.5: Average temperatures, self-diffusion coefficients, static dielectric constants, and Debye relaxation times of liquid water.

Methods	T_{NVE} [K]	D_{PBC} [Å^2/ps]	ϵ	τ_D [ps]
Experiment		0.18[a]	78.4[b]	8.5[c]
AM1	305 ± 3	0.46	31.7 ± 2.0	1.2
AM1-W	310 ± 3	0.01	35.1 ± 9.0	200
AM1PG-W	302 ± 3	0.83	116.6 ± 6.6	1.3
AM1-LW (this work)	298 ± 2	0.37	95.4 ± 9.6	6.7

a. Experimental self-diffusion coefficient (0.23 Å^2/ps[169]) minus the finite size correction[168] that would apply to a system of 64 molecules (0.05 Å^2/ps)

b. Reference [170]

c. Reference [171]

3.4.6 Static and dynamic dielectric constants

The dielectric constant of the liquid, ϵ, is calculated using the following formula:[172]

$$\epsilon = \epsilon_\infty + \frac{4\pi}{3Vk_BT}\frac{(\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2)}{\varepsilon_0} \tag{3-5}$$

where $\epsilon_\infty = 1$ is the dielectric constant of vacuum, $\mathbf{M}$ is the total dipole moment of the simulation box, V is the volume of the simulation box, T is the temperature of the system, k_B is Boltzmann constant, and ε_0 is the permittivity of vacuum. The static dielectric constants obtained from Eq. (3-5) using different models are shown in Table 3.5. As dielectric constant grows with the amplitude of the dipole moment fluctuations, the models, which yield correct molecular dipole moment of water, tend to estimate the correct dielectric constant. The dielectric constant versus molecular dipole moment curve shows about 50% difference in dipole moments can lead as much as six times differences in the dielectric constant.[173] The AM1 gives the correct gas-phase dipole moment of water monomer, but yields low dielectric constant of water, which is

likely due to the underestimation of molecular dipole moment of water in liquid. Despite yielding a larger and more realistic number of hydrogen bonds, the AM1-W model underestimates the dielectric constant, going against the generally expected trend.[174] AM1PG-W and AM1-LW models show significant improvement on reproducing dielectric constant over other models. These improvements are partly due to their performances on yielding dipole moment of water in liquid. For instance, the AM1-LW model slightly overestimates the dielectric constant with value 95.4 in comparison to experimental value of 78.4. The overestimation may arise from the slightly overestimation on dipole moments of monomer in liquid water. Figure 3.5 shows the running average dielectric constants with errors as a function of simulation time. It is observed that the convergence of dielectric constant is very slow for all models except AM1.

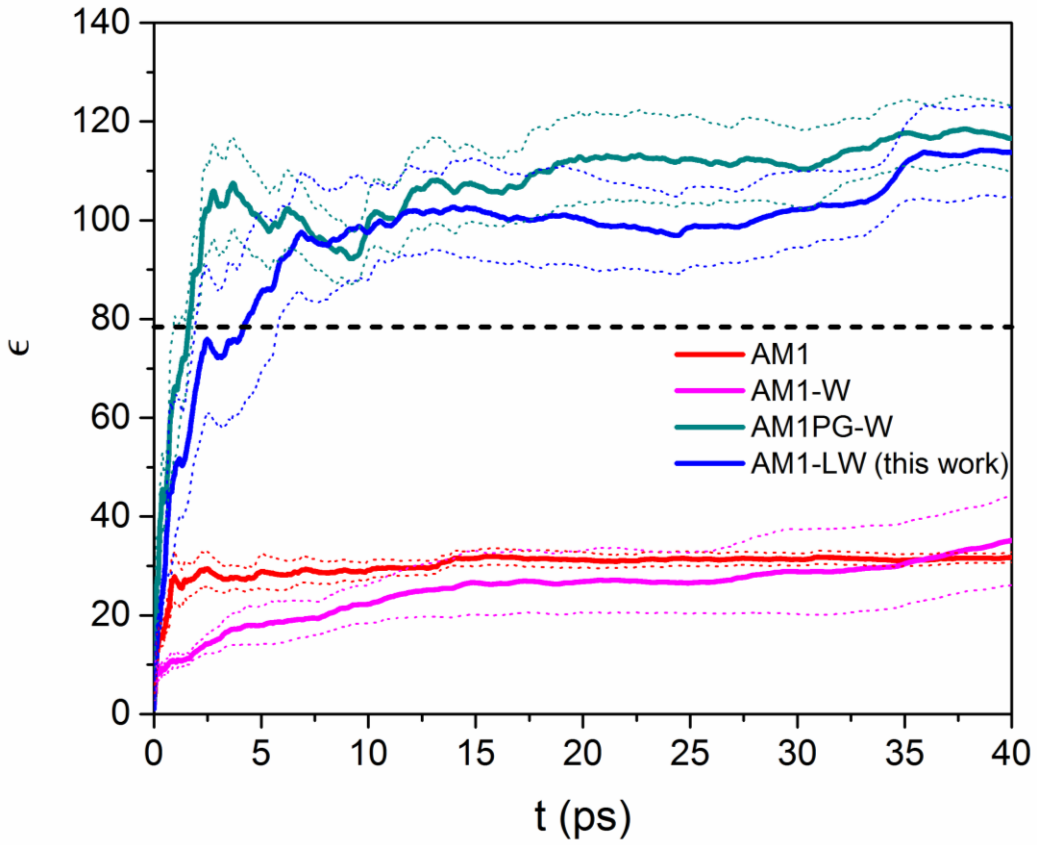


Figure 3.5: Convergence of the estimated static dielectric constant as a function of time as $\mathbf{M}(t)$ is accumulated. Horizontal dashed line shows the experimental value. Dotted lines are the running estimated standard errors from independent multiple simulations (see Section 3.3.2 for details).

Frequency dependent dielectric constant draws some insight of relaxation dynamics of the liquid and dipole moment orientational dynamics. Frequency dependent dielectric constant is calculated using the following relation:[175–177]

$$\frac{\epsilon(\omega) - \epsilon_\infty}{\epsilon_0 - \epsilon_\infty} = 1 - i\omega\mathcal{L}_{i\omega}[\phi(\tau)] \tag{3-6}$$

where $\mathcal{L}_{i\omega}$ denotes the Laplace transform of the normalized autocorrelation function $\phi(\tau)$ of the system's total dipole moment. $\phi(\tau)$ is defined as following:

$$\phi(\tau) = \frac{\langle \mathbf{M}(\tau) \cdot \mathbf{M}(0) \rangle}{\langle |\mathbf{M}(0)|^2 \rangle} \tag{3-7}$$

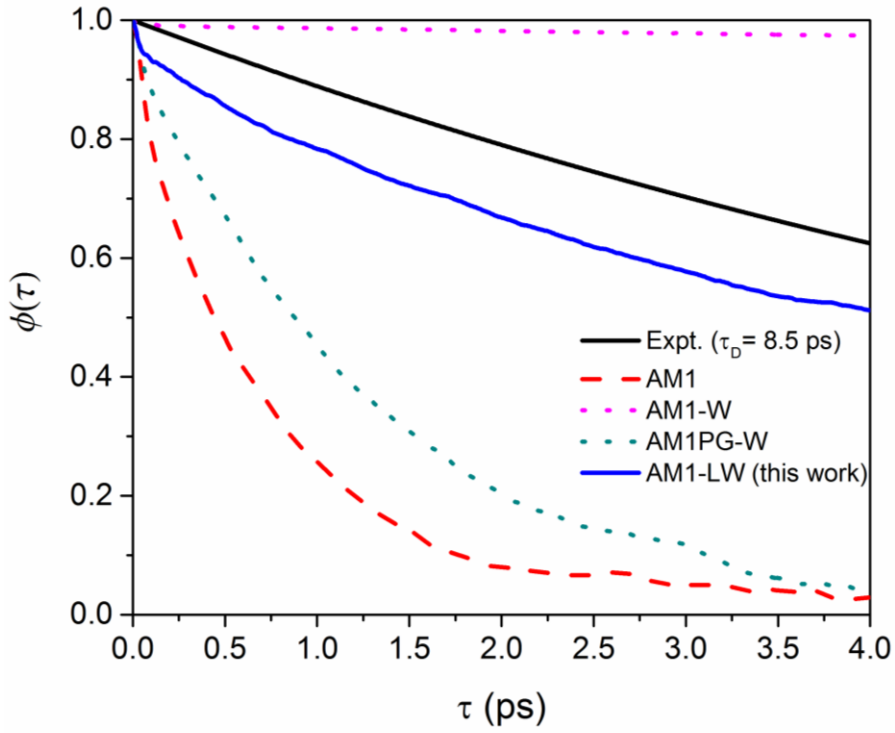


Figure 3.6: Time autocorrelation function $\phi(\tau)$ of total dipole moment of the system. Black line represents a single exponential decay using the experimental Debye relaxation time τ_D=8.5 ps [171].

Figure 3.6 shows the autocorrelation function of total dipole moment of the system calculated from Eq. (3-7). $\phi(\tau)$ decays quickly for AM1 and AM1PG-W but very slowly for

AM1-W. The slow decay of $\phi(\tau)$ for AM1-W is consistent with the fact that the model has almost no diffusivity (see Table 3.5). Since $\phi(\tau)$ decays exponentially, Debye relaxation time τ_D, which is a key parameter for describing solvation dynamics,[178] is calculated from the long range exponential fit of $\phi(\tau)$ as[175]

$$\phi(\tau) = A\, e^{-\tau/\tau_D} \tag{3-8}$$

The shape of $\phi(\tau)$ based on experimental Debye relaxation time (τ_D = 8.5 ps) is also presented in Figure 3.6. As shown in Figure 3.6, the AM1-LW model slightly underestimates the Debye relaxation time as compared to experimental value. The AM1-LW model also captures the signature of the fast damped oscillations and the second relaxation process. The existence of these two-relaxation modes in liquid water is consistent with previous study on dielectric relaxation in liquid water.[179] The calculated Debye relaxation times are presented in Table 3.5. It is found that AM1 and AM1PG-W yield very low Debye relaxation times of 1.2 and 1.3 ps respectively and AM1-W gives very large value of 200 ps. Unlike other models, the AM1-LW model yields a Debye relaxation time of 6.7 ps in good agreement with the experimental value of 8.5 ps.

The dynamics of the liquid can be studied from the frequency-dependent dielectric constant. The frequency-dependent dielectric constant is written as $\epsilon(\omega) = \epsilon'(\omega) - i\epsilon''(\omega)$, where the real part $\epsilon'(\omega)$ describes dispersion and the imaginary $\epsilon''(\omega)$ describes absorption characteristics of the liquid.[180] More importantly, the complex part describes the radiative transfer of microwave radiation in water. Figure 3.7 presents frequency dependent dielectric constant of water from the AM1-LW. Despite the overestimation of the static dielectric constant, the AM1-LW model captures the overall features of both real and imaginary dielectric constants as a function of frequency. More importantly, it reproduces imaginary part at low frequency range, which is crucial for describing absorption spectra.

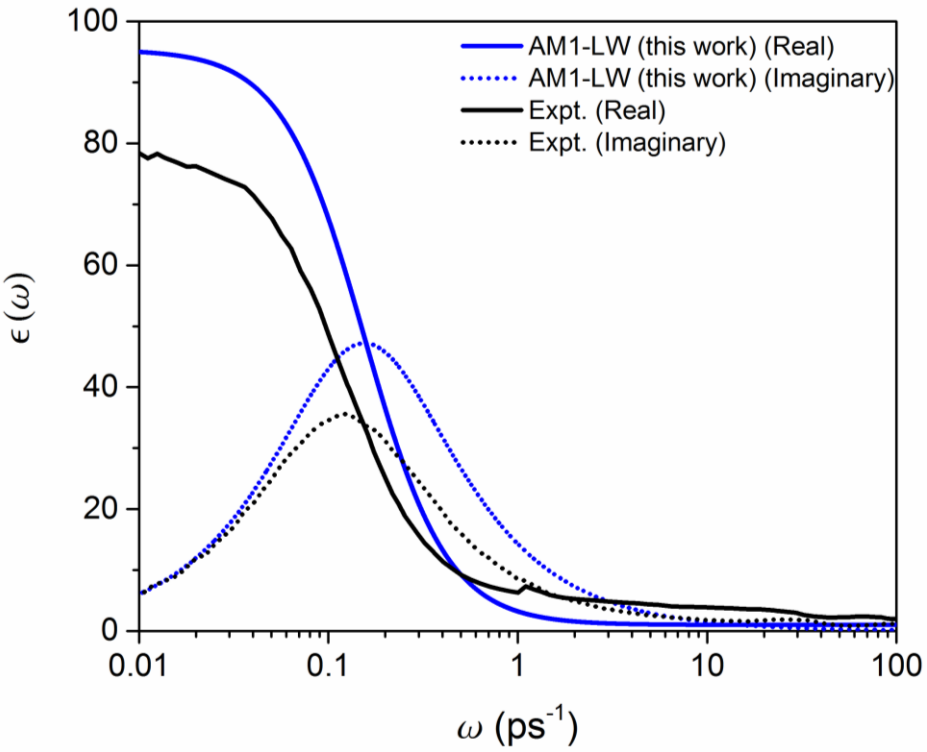


Figure 3.7: Real (solid) and imaginary (dotted) parts of frequency-dependent dielectric constant for water obtained from the AM1-LW model (this work) using Eq. (3-6), compared to experiment. Experimental curves are taken from ref. [177].

3.4.7 Harmonic vibrational frequencies and infrared spectra

Table 3.6 presents vibrational frequencies of the water monomer and dimer calculated for different SE models. Experimental/ab initio values are also presented for comparison. In the case of the dimer, vibrational modes are presented for both proton-acceptor (A) and proton-donor (D) water molecules. Compared to experimental vibrational frequencies, AM1 systematically underestimates the frequencies of symmetric (*sym*) and anti-symmetric (*anti*) modes and overestimates the frequencies of bending modes indicating bonds are too flexible and angle is too stiff. AM1-W systematically overestimates the frequencies of both symmetric and anti-symmetric modes and underestimates bending mode while AM1PG-W underestimates all frequencies. Even though vibrational frequencies were not explicitly adjusted during the parameterization of AM1-LW, as shown in Table 3.6, the AM1-LW model notably improves the accuracy of vibrational

frequencies for both water monomer and dimer in comparison with reported experimental/ab initio results.

Table 3.6: Vibrational frequencies of water monomer and dimer predicted from different SE models. Experimental/ab initio values are taken from reference [181].

Methods	**Monomer**					**Dimer**			
	ν_1(*sym*) cm^{-1}	ν_2(*bend*) cm^{-1}	ν_3(*anti*) cm^{-1}	$\Delta\nu_{13}$ cm^{-1}	$\Delta\nu_{rms}$ cm^{-1}	ν_1(*sym*) cm^{-1}	ν_2(*bend*) cm^{-1}	ν_3(*anti*) cm^{-1}	$\Delta\nu_{rms}$ cm^{-1}
Expt./ ab initio	3832.2	1648.5	3942.5	110.3		3797 (A) 3718 (D)	1669 (A) 1653 (D)	3899 (A) 3881 (D)	
AM1	3518.4	1879.1	3588.7	70.3	303.8	3490 (A) 3489 (D)	1878 (A) 1900 (D)	3575 (A) 3574 (D)	284.5 (A) 263.1 (D)
AM1-W	3950.2	1312.9	4055.3	105.1	215.5	3856(A) 3489(D)	1387 (A) 1458 (D)	3986 (A) 4084 (D)	173.8 (A) 209.5 (D)
AM1PG-W	3669.2	1515.1	3754.3	85.1	163.1	3597 (A) 3400 (D)	1513 (A) 1547 (D)	3699 (A) 3773 (D)	186.5 (A) 203.3 (D)
AM1-LW (this work)	3921.1	1492.9	4024.9	103.8	113.9	3857 (A) 3610 (D)	1527 (A) 1552 (D)	3979(A) 4032 (D)	100.3 (A) 122.0 (D)

Table 3.6 also reports $\Delta\nu_{13} = \nu_3 - \nu_1$, the frequency shift between the symmetric and antisymmetric stretching modes of the water monomer. The models with more flexible bonds (AM1 and AM1PG-W) underestimate the frequencies shift, while the models with more rigid bonds (AM1-W, AM1-LW) AM1 and AM1PG-W models yield symmetric and antisymmetric frequency bands, which are unrealistically close to each other (small $\Delta\nu_{13}$). Even though $\Delta\nu_{13}$ from AM1-W is closer to experimental value, both symmetric and anti-symmetric frequencies are highly overestimated. Compared to experimental value (110.3 cm^{-1}), the AM1-LW model (103.8 cm^{-1}) gives agreeable value of $\Delta\nu_{13}$. The root mean square deviations $\Delta\nu_{rms}$ of all three vibrational frequencies ν_1, ν_2, and ν_3 are also calculated from four different SE models. As shown

in Table 3.6, the AM1-LW model gives lowest $\Delta\nu_{rms}$ of 113.9 cm^{-1} for water monomer, and 103.3 cm^{-1} (acceptor) and 122.0 cm^{-1} (donor) for water dimer. The accuracy of the models for vibrational frequencies increases in the order AM1 < AM1-W < AM1PG-W < AM1-LW.

The dynamical properties of liquid water are further investigated by calculating the infrared (IR) absorption spectrum. The IR absorption coefficient is calculated from the Fourier transform of the autocorrelation function of the total dipole moment of the system. The product of absorption coefficient $\alpha(\omega)$ and refractive index $n(\omega)$ can be written as[97,99]

$$\alpha(\omega)n(\omega) = \frac{2\pi\omega^2\beta}{3cV} \times \int_{-\infty}^{+\infty} dt\, \langle \mathbf{M}(t) \cdot \mathbf{M}(0) \rangle e^{-i\omega t} \tag{3-9}$$

where $\beta = 1/k_B T$ (k_B and T are the Boltzmann constant and temperature), V is the volume of the system, and c is the speed of light in vacuum. The quantum time-correlation function is calculated from the classical function by applying a harmonic quantum correction factor Q_{HA} of the form:

$$Q_{HA} = \frac{\beta\hbar\omega}{1 - e^{-\beta\hbar\omega}} \tag{3-10}$$

It was found that a harmonic quantum correction factor performs better than the Kubo[182] and the Schofield[183] corrections in capturing IR spectra of water. As previously reported,[97] harmonic correction factor satisfies the fluctuation-dissipation theorem and restores the detailed balance condition.

IR spectra of liquid water calculated from AM1, AM1-W, AM1PG-W, and AM1-LW models are shown in Figure 3.8 along with experimental spectrum. AM1, AM1-W, and AM1PG-W models do not reproduce the experimental IR spectra and hence they are not discussed further. Unlike other models, AM1-LW yields a nature of the spectrum with good agreement to experiment. The shape of the spectrum regardless of intensity is similar to experiment especially in the low frequency region below 1000 cm^{-1} that corresponds to librational modes. The positions of libration, HOH bending and OH stretch modes are approximately 760, 1590 and 3390 cm^{-1} respectively which are within 100 cm^{-1} of experimental values of ~650, ~1650, ~3400 cm^{-1}.[184] Previous studies show that OH stretch frequency of liquid water is highly correlated with O–O

distances and intermolecular O–H distances between hydrogen-bond donor and acceptor water molecules.[185–187] AM1-LW results for the diffusivity and vibrational frequency shows the model captures the translational, rotational, and vibrational motions of the liquid water.

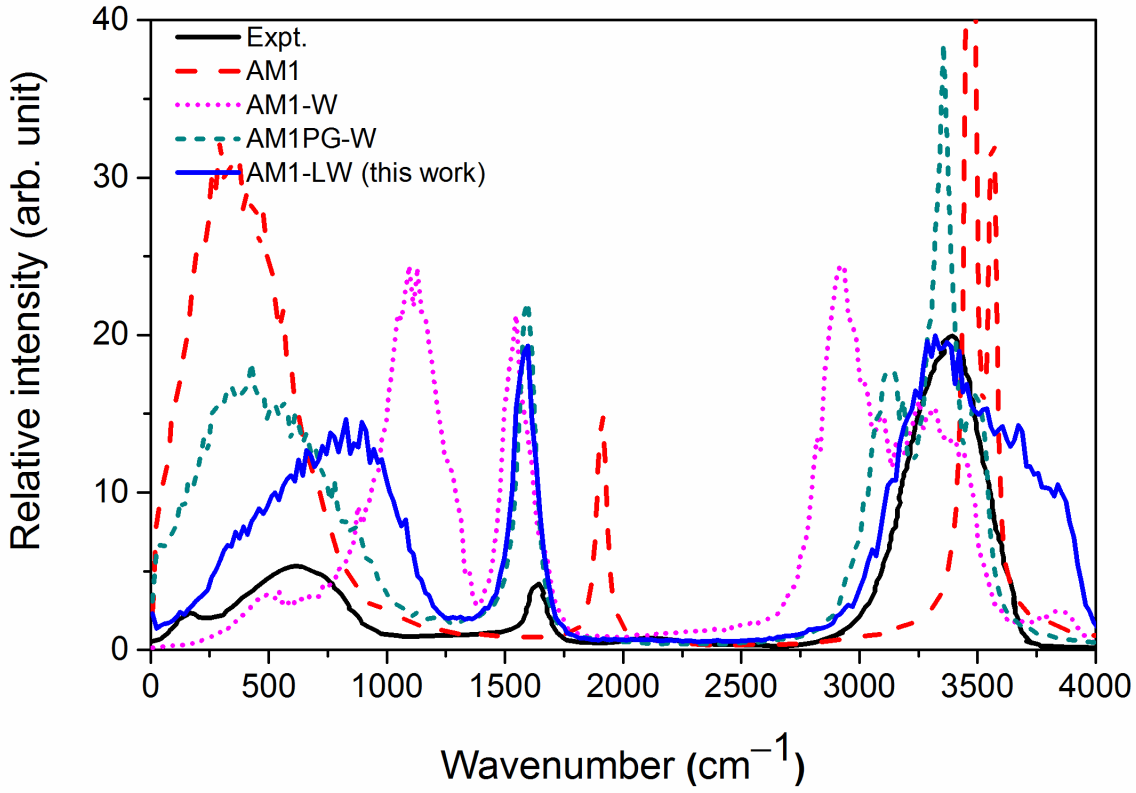


Figure 3.8: IR spectrum (arbitrary units) of liquid water obtained from AM1, AM1-W, AM1PG-W and AM1-LW models in comparison with experiment[188,189].

3.5 Conclusion

A re-parameterization of AM1 model, called AM1-LW, has been presented for describing gas-phase proton transfer reactions of water and structure of liquid water. Various gas-phase properties from the AM1-LW model were compared with standard semiempirical models (AM1, PM3, and RM1). The AM1-LW model shows a great improvement over standard models in reproducing a number of gas-phase properties of water, including proton transfer related quantities.

From the molecular dynamics simulations, it is found that the AM1-LW model provides improved accuracy for various static and dynamical properties of liquid water over AM1, AM1-W, and AM1PG-W models. In particular, the AM1-LW model better reproduces the experimentally observed structure of liquid water. The AM1-LW model yields O–O coordination and enthalpy of vaporization in agreement with experiment, which reflects the balance of polarization and dispersion interactions. Sensitive dynamical properties such as self-diffusion coefficient, dielectric constant, and Debye relaxation time were also investigated. Compared to other models, AM1-LW shows a great improvement on reproducing some of the properties of water. Interestingly, the AM1-LW model reproduces the imaginary part of dielectric constant at lower frequency range, which leads to infrared spectrum line shapes of water in good agreement with experimental spectrum.

The results from this work confirm that the poor performance of the SE models can be improved without modifying the original Hamiltonian, by adopting a powerful enough parameterization technique and minimizing a more specific error function. A recent study from Welborn *et al.*[147] also suggests that properly parameterized SE methods can yield correct structural and electronic properties for water in both gas and liquid phases and concluded that the poor descriptions from NDDO based SE methods are coming from the poor parameterization strategies.[147] Despite having a few limitations, the AM1-LW model shows marked improvement over other SE models. Hence, AM1-LW is more suitable to the study of condensed-phase properties of water and represents an excellent compromise between computational cost and the accuracy. Future work will be aimed at developing reaction-specific semiempirical models to describe the structural and chemical properties of biological systems in which water plays an essential role.

4 Development of semiempirical models for zinc-water and zinc-cysteine binding in metalloproteins

4.1 Abstract

Zinc plays crucial roles in biocatalysis and in stabilizing protein structure. In principle, zinc-catalyzed enzymatic reactions and zinc binding in proteins can be modeled using high-level quantum methods. However, due to their very high computational cost, these methods are prohibitive to the study of zinc metalloproteins. Quantum mechanical semiempirical methods based on the neglect of diatomic differential overlap approximation introduce empirical parameters for many expensive integrals in ab initio methods, which make calculations fast while retaining acceptable accuracy. We aim to develop a semiempirical model for Zn that correctly describes the hydration structure of Zn^{2+} and the metal-induced pK_a shifts for ligands in Zn^{2+} complexes mimicking the active site of some zinc-metalloproteins. The standard AM1 semiempirical model is re-parameterized to reproduce structural properties and energetics (complexation energies, ligand-exchange energies, proton transfer energies, and proton affinities) of a set of compounds (the training set) calculated from hybrid density functional theory [B3LYP/6-311++G(d,p)]. The training set contains $Zn^{2+}\cdot H_2O$, $Zn^{2+}\cdot H_2S$ (model for cysteine), and H_2O/H_2S ligand exchange reactions in hexa- and tetra-coordinated Zn^{2+} complexes, in both their neutral and deprotonated forms. Semiempirical parameters for Zn and S are optimized using a genetic algorithm approach. Parameters for H and O are used from our previous study on proton transfer in water (see Chapter 3). The transferability of the model is assessed for a large set of $Zn^{2+}\cdot(H_2O)_n\cdot(H_2S)_m$ complexes (with $n + m = 4$, 5 or 6) (the testing set). The optimized model provides a significant improvement over standard AM1 and PM3 models for both training and testing sets. For instance, the average unsigned error in reaction energies from the optimized model is 2.27 kcal/mol compared to 30.30 (AM1) and 22.30 (PM3) kcal/mol in the training set and 4.02 kcal/mol compared to 24.29 (AM1) and 48.18 (PM3) kcal/mol in the testing set. The optimized model is used to simulate the hydration structure of Zn^{2+} and it yields correct hexacoordinated hydration structure with average Zn···O distance of 2.13 Å in good agreement with the experimental value of 2.06 Å. The model is further applied to study the metal-induced pK_a shift of water and hydrogen sulfide in liquid water. The optimized model yields pK_a of zinc-

bound water of 9.0 and zinc-bound hydrogen sulfide of 2.0 in excellent agreement with available experimental/theoretical results. The optimized model can be a reliable semiempirical model to study different structural and catalytic mechanisms at the active sites of zinc metalloproteins.

4.2 Introduction

Metal ions are important for their biological functions of metalloproteins. They activate or regulate the function of proteins by causing change in structure or sewing catalytic center and are also involved in protein stabilization, protein-protein interactions, and protein-ligand binding.[57,190–192] Zinc is the second most abundant metal in biological systems, after iron. It plays crucial roles in biocatalysis and in stabilizing protein structures. Zn also involves in regulatory and transport mechanisms.[193] For catalytic roles, zinc is often selected over other divalent ions because of its flexible coordination geometry, fast ligand exchange (with a water exchange rate around 10^7-10^8 sec^{-1}),[194] Lewis acidity, and lack of redox activity.[43]

Upon binding to the ligands, Zn^{2+} increases ligands' acidity. Zn-bound water plays crucial roles in the catalytic sites of Zn^{2+} proteins. Generally, Zn^{2+} binds one water and three amino acids in catalytic active sites but binds four amino acids in structural active sites.[195] . Common binding motifs of Zn^{2+} enzymes at catalytic sites include His/His/His/H_2O (e.g. metallo-β-lactamase, carbonic anhydrase), and His/Asp/Asp/H_2O (e.g. histone deacetylases) and structural sites include Cys/Cys/Cys/His (e.g. nucleocapsid p7), Cys/Cys/Cys/Cys (e.g. breast cancer-associated protein 2).[196] A recent survey on the coordination environment of zinc showed that cysteine and histidine account for ~72% and ~25% of total coordinating ligands respectively.[192] While Zn^{2+} normally binds four ligands in a tetrahedral geometry in protein active sites,[197,198] it has the ability to form higher (up to 6) coordination geometries in few proteins.[199,200] Understanding the hydration structure of Zn^{2+}, the change in water acidity upon binding to Zn^{2+}, and the effect of additional ligands on the pK_a shift of water have particular interest to protein chemistry. Study of Zn^{2+} binding water also helps for gaining insight into binding features of Zn^{2+}-ligand complexes and biological importance in zinc metalloproteins.

In proteins, Zn^{2+} binds cysteine preferably to any other amino acid. The ionization state of the functional thiol group of cysteine is highly sensitive to the local protein environment because its pK_a value (7.4) is closed to physiological pH.[193] The interaction of cysteinate R–S^- lone pair

with Zn^{2+} has a particular interest as it determines the nature of the interaction (e.g. electrostatic or covalent or ionic) and the soft characteristics of the ligand in zinc metalloenzymes.[201] To understand the zinc-ligand interactions analogous to metal binding in protein, very high-level quantum mechanical methods are often used for small biomolecular clusters mimicking the functional groups found in zinc metalloprteins.[202–204]

State-of-the-art studies of the catalytic roles of Zn^{2+} in proteins uses hybrid quantum mechanics/molecular mechanics (QM/MM) approaches, where the active site of the protein is described with a quantum mechanical theory and the rest of the system is defined with classical mechanics. While classical mechanical force-field methods allow for studying the structural dynamics of proteins, they typically do not allow bond breaking and cannot be used to study reaction mechanisms. That is, force fields cannot be used to study reactions at the catalytic site of the enzymes. Moreover, current widely used force-fields CHARMM,[205] AMBER,[206] and OPLS[207,208] are designed for the protein residues, not for the metal centered reaction centers. Recently new force fields are being developed to study solvation structure of Zn^{2+} in water[209] and zinc-ligand interactions (binding coordination energies and zinc-ligand binding distances) in metalloproteins[198,210–214]. Because of the inherited limitations of force field methods, these methods are only applicable to study hydration free energy and structural dynamics of ions in a solvent. Semiempirical (SE) methods are based on the Hartree–Fock formalism, but use approximations to avoid computationally expensive steps and introduce empirical parameters fitted from empirical data. SE methods tend to reproduce correct structures and energetics of the biomolecules.[118] Thus, SE methods provide a compromise between the computational costs of ab initio quantum mechanical calculations and the electrons-related limitations of force fields.

Standard SE models based on the neglect of diatomic differential overlap (NDDO) approximations[72] such as AM1[73] and PM3[86] were developed to reproduce the structural and thermochemical properties of small organic compounds, however, they inadequately describe zinc-ligands (for example, Zn···O or Zn···S) interactions.[131,204,215] To improve the accuracy of the SE models, reaction-specific re-parameterization of original SE models are often carried out.[93,133,216–219] In this work, we optimize the standard AM1[73] SE model for zinc metalloproteins. The parameters for Zn and S are optimized to reproduce the structural properties and energetics (complexation energies, ligand exchange energies, proton affinities, and proton transfer energies)

of zinc-complexes that mimic the active sites of zinc metalloproteins. For hydrogen and oxygen, previously optimized AM1 parameters, to study proton transfer in water, are used.[220] The optimized model is then used to explore the solvation structures of zinc ion in liquid water and zinc ion in liquid hydrogen sulfide and the influence of zinc-coordination on ligands' pK_a.

4.3 Computational Methods

Re-parameterization of SE models was done in three steps. First, the training set of a small set of model compounds to represent the active site of the zinc metalloproteins was prepared. Second, the parameterization procedure was carried out in a similar manner as reported in the previous work[120,220] using the training set. Third, the transferability of the optimized model was assessed by applying on the large set of compounds named as the testing set.

Finally, the optimized model was subjected to gas- and liquid- phases simulations on zinc binding ligands. The performance of the optimized model was compared with original AM1[73] and PM3[86] models for both gas- and liquid-phase results along with available theoretical and experimental results.

4.3.1 Preparation of training set and gas phase calculations

Optimization of the AM1 parameters for Zn and S was performed based on the gas-phase properties of molecular complexes mimicking the active site of zinc proteins. The training set contains structural and energetic properties associated with biomimetic complexes of Zn^{2+}, H_2O/OH^-, and H_2S/HS^- (models for neutral and deprotonated cysteine) and H_2O/H_2S ligand exchange reactions in hexa- and tetra-coordinated Zn^{2+} complexes. In addition, neutral and deprotonated $H_2S\ldots H_2O$ binary complexes and proton transfer profiles in $H_2S\ldots H_2O$ were also included in the training set. The compounds and associated reactions used in the training set during the parameterization are presented in Figure 4.1.

Gas-phase DFT calculations were performed on B3LYP/6-311++G(d,p) level of theory using Gaussian 09[150] and semiempirical calculations were done using MOPAC 07.[151] During geometry optimization in MOPAC 07, the termination criteria GNORM was set to be 1 kcal/mol/Å during the parameterization and was set to be 0.01 kcal/mol/Å for the final

assessment of the SE models. The final optimized coordinates from DFT were used as the initial coordinates for SE geometry optimization.

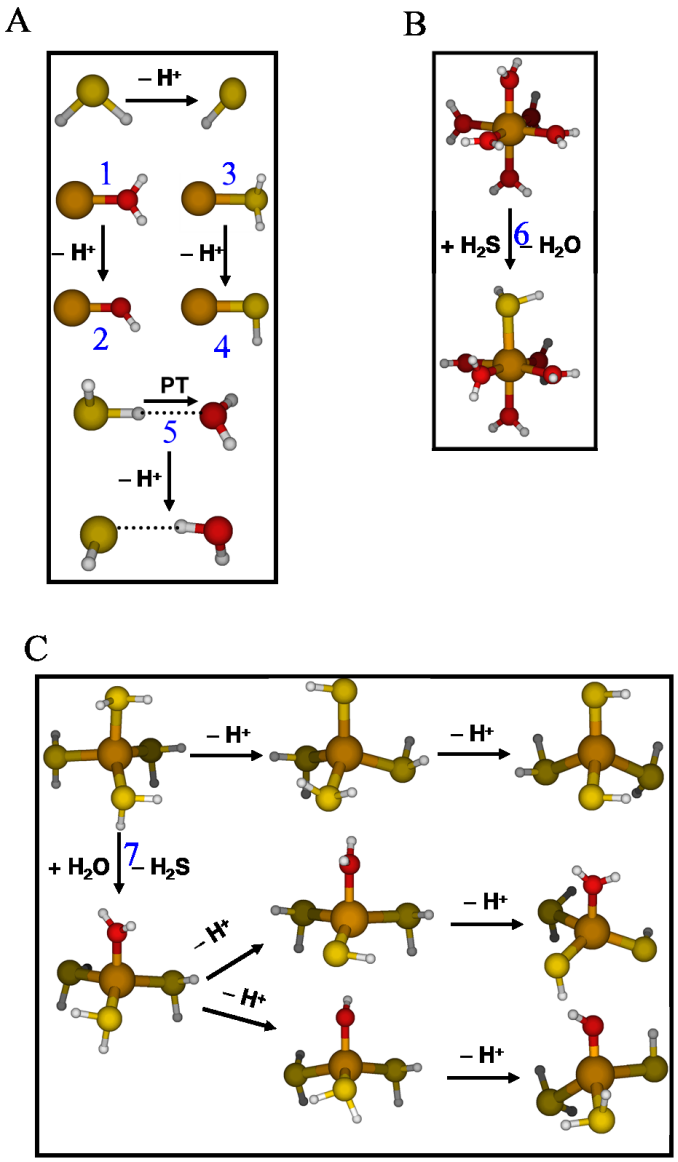


Figure 4.1: Compounds and reactions used in the training set. (A) Complexation of simple binary complexes in their neutral and deprotonated forms (B) Ligand exchange reaction in hexa-coordinated zinc complexes and (C) Tetra-coordinated zinc complexes in their neutral and deprotonated forms including ligand exchange reaction in the neutral form. The numbers (1-5) identify the complexes used for complexation energies. Deprotonation reactions (labeled "$–H^+$"), proton transfer reactions (labeled "PT"), and ligand exchange energies (shown by numbers 6 and 7)

Since proton transfer mechanisms play crucial roles in chemical and biomolecular processes in the living organism, it is important to have a model, which also reproduces proton transfer profiles. One way to incorporate this is by reproducing the proton affinity of the model compounds of the active sites of the proteins. The proton affinity, E_{PA}, was calculated as following:

$$E_{\mathrm{PA}}^{\mathrm{DFT}} = -\Delta E - ZPE + \frac{5}{2}RT$$
$$E_{\mathrm{PA}}^{\mathrm{SE}} = -\Delta H + H_{\mathrm{f}}(\mathrm{H}^+) \tag{4-1}$$

where E is electronic energy, ZPE is zero-point correction energy, R is gas constant, and T is temperature. ΔE and ΔH are the differences in electronic energy and heat of formation between the protonated and deprotonated forms of the complex. For SE models, experimental heat of formation of proton [$H_{\mathrm{f}}(\mathrm{H}^+)$= 367.2 kcal/mol][221] was used to calculate proton affinity. Proton transfer profiles were calculated by scanning the distance between H(donor) and O(acceptor) in $H_2S \cdots H_2O$ complex (going from $H_2S \cdots H_2O$ to $HS^- \cdots H_3O^+$) from 2.0 to 1.0 Å in the steps of –0.25 Å while keeping S(donor), H(donor), and O(acceptor) collinear. Five points (each represents as the k^{th} constraint) along the proton transfer energy profiles were calculated, from the energy difference between k^{th} constraint complex [$E(k)$] along the proton transfer energy scan and the energy of the complex for H(donor)⋯ O(acceptor) distance of 2.0 Å [$E(0)$] as:

$$E_{\mathrm{PT}}(k) = E(k) - E(0) \tag{4-2}$$

4.3.2 Error function and parameterization procedure

The SE parameters for Zn and S were obtained by minimizing the error function given by

$$\begin{aligned}\chi &= W_{\mathrm{S}} \sum_i |MSD_i| + W_{\mathrm{B}} \sum_j \left|B^{\mathrm{model}}(j) - B^{\mathrm{ref}}(j)\right|^2 \\ &+ W_{\mathrm{R}} \sum_i \left|E_{\mathrm{R}}^{\mathrm{model}}(i) - E_{\mathrm{R}}^{\mathrm{ref}}(i)\right| + W_{\mathrm{PA}} \sum_i \left|E_{\mathrm{PA}}^{\mathrm{model}}(i) - E_{\mathrm{PA}}^{\mathrm{ref}}(i)\right| \\ &+ W_{\mathrm{PT}} \sum_i \left[\frac{1}{n_{\mathrm{PT}}} \sum_{k=1}^{n_{\mathrm{PT}}} \left|E_{\mathrm{PT}}^{\mathrm{model}}(i,k) - E_{\mathrm{PT}}^{\mathrm{ref}}(i,k)\right|\right]\end{aligned} \tag{4-3}$$

where W_S (10^3 kcal/mol/Å^2) is a weighting factor for mean square deviations (MSD) in molecular structures, W_B (10^4 kcal/mol/Å^2) is a weighting factor for bond lengths (B), W_R (50) is a weighting factor for reaction energies (E_R), W_{PA} (50) is a weighting factor for proton affinities (E_{PA}), and W_{PT}(50) is a weighting factor for proton transfer energies (E_{PT}). The units of weighting factors were chosen in a way that the value of the error function is in kcal/mol. $n_{PT} = 5$ is the number of points along each proton transfer profile. Superscripts "model" and "ref" represent SE and DFT results. $j = 8$ represents the number of Zn···O/S bonds in [Zn^{2+}·H_2O] (one bond), [Zn^{2+}·H_2S] (one bond), and [Zn^{2+}·$6H_2O$] (six bonds) complexes. i represents complex index and k (= 1 to 5) represents each constraint point on the PT scan. The error function contains 17 mean square deviations in structures, 8 Zn···O/S bonds, 10 proton affinities, 5 proton transfer energies, and 7 reaction/complexation energies (see Figure 4.1).

12 parameters for Zn were first optimized followed by the 21 parameters for S. Original AM1[73] parameters for both Zn[222] and S[223] were used as initial parameters and allowed ±50% change from their original values during the optimization. Previously optimized H and O parameters for proton transfer reactions in water were used for the elements hydrogen and oxygen.[220] A parallel version of PIKAIA, a biologically inspired genetic algorithm with the evolution of 100 individuals for 300 generations, was used to optimize the set of SE parameters. Once PIKAIA was started, fitness was evaluated for all chromosomes (here chromosomes are SE parameters) in the population and parameters set with maximum fitness score was collected. New generations were obtained by the genetic crossover of the previous pair of individuals, followed by random mutation. Three rounds of parameterization were needed to converge the error value. The parameterization was repeated by reducing the search region to only ±20%, ±10%, ±5% and ±1% change from previously optimized parameters for each element. Only two rounds needed to decrease the error values for each reduced search space. Finally, no improvement on the error value was found and the parameterization was stopped and the final optimized parameters for zinc and sulfur were collected.

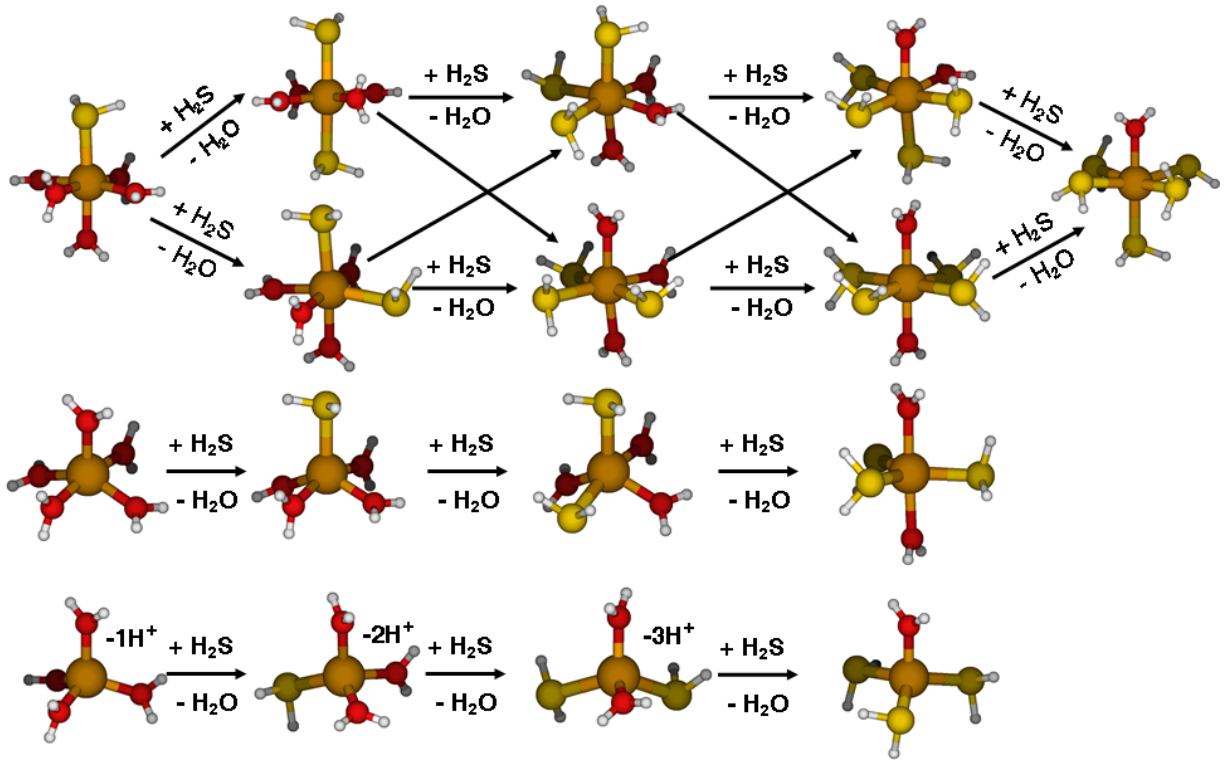


Figure 4.2: Testing set. Arrows show the ligand exchange reactions where hydrogen sulfide replaces zinc-bound water one at a time from hexa-, penta-, and tetra-coordinated complexes. $-1H^+$ (one from H_2O), $-2H^+$ (one from H_2O, one from H_2S), and $-3H^+$ (two from two H_2O, one from H_2S; two from two H_2S deprotonation results H_2O leaving first coordination shell) indicate one, two and three deprotonated structures associated with the shown parent structures.

The optimized model was tested for its transferability on reproducing structural and energetics of associated to the complexes that composed of hexa-, penta-, and tetra-coordinated binary and tertiary zinc complexes with H_2O and H_2S (see Figure 4.2). Three classes of ligand exchange reactions were considered as shown in Figure 4.2. First, ligand exchange reactions in hexacoordinated zinc-bound complexes where each H_2O was replaced by one H_2S at a time. All possible, *cis*, *trans, facial,* and *meridional*, structural configurations were included. Second, similar ligand exchanges in penta- and tetra-coordinated zinc complexes were also included. Six proton affinities (one in $Zn^{2+}{\cdot}4H_2O$, two in $Zn^{2+}{\cdot}3H_2O{\cdot}1H_2S$, and three in $Zn^{2+}{\cdot}2H_2O{\cdot}2H_2S$) were computed (see Figure 4.2). During a geometry optimization, it is observed that some of the ligands leave the coordination with zinc. Only complexes, for which all ligands are stable in the first coordination shell of zinc, were considered. $Zn^{2+}{\cdot}6H_2S$ and $Zn^{2+}{\cdot}5H_2S$ structures do not form

stable structures, and thus they were discarded. 20 mean square deviations in structures, 6 proton affinities, and 18 ligand exchange reactions between water and hydrogen sulfide were included in the testing set.

4.3.3 Molecular dynamics simulations

Molecular dynamics (MD) simulations were performed with the CP2K package[152] using AM1, PM3, and the optimized model. Four systems are simulated using each model; Zn^{2+} in 63 water molecules, Zn^{2+}-H_2S and Zn^{2+}-SH^- solvated in 62 water molecules, and Zn^{2+} solvated in 63 hydrogen sulfide molecules. All simulations were performed with the canonical ensemble (*NVT*) using a cubic box of 12.38 Å of side (16.75 Å of side for zinc solvated in liquid H_2S) with periodic boundary conditions. Each system was simulated for 40-ps (10-ps equilibration and 30-ps production) with a time step of 0.5 fs. The simulations were performed at T = 300 K, controlled using a massive Nosé-Hoover thermostat with a chain length of 3 and time constant of 2000 fs.

4.4 Results and discussion

The optimized parameters for S and Zn are presented in Table 4.1 along with parameters for H and O from previous work.[220]

4.4.1 Performance on gas-phase clusters

Figure 4.3 (panels a to e) shows, in comparison to DFT results, the performance of AM1, PM3 and the optimized AM1 model on Zn···O/S bond lengths, mean square deviations in the overall geometries, proton affinities, reaction energies, and proton-transfer energy profiles of the training set complexes. As expected, the optimized model has better correlation with DFT results than AM1 and PM3 (see Table 4.2). Although the optimized model underestimates the Zn-S bond length (Figure 4.3 (a)), it yields a small average unsigned error (AUE) in bonds of 0.03 Å, compared to 0.06 Å for AM1 and 0.07 Å for PM3. The model also shows AUEs on proton affinities and reaction energies of 4.53 and 2.27 kcal/mol, compared to 15.13 and 30.30 kcal/mol for AM1 and 7.86 and 22.30 kcal/mol for PM3.

Table 4.1: Optimized AM1 parameters for S and Zn. Parameters for H and O obtained from our previous work are also presented.[220] Symbols have their usual meanings as found in reference [73].

Parameters	**H**	**O**	**S**	**Zn**
U_{ss} (eV)	−12.560002	−124.312581	−45.503774	−18.409395
U_{pp} (eV)		−79.588146	−48.725822	−14.386244
ζ_s (au)	1.078537	3.497468	2.045720	2.170903
ζ_p (au)		2.528886	1.826787	1.320255
β_s (eV)	−4.860006	−37.859064	−4.854272	−1.249800
β_p (eV)		−37.180813	−7.686882	−5.141664
G_{ss} (eV)	14.331944	21.583836	16.376114	8.305369
G_{sp} (eV)		18.193541	5.011686	8.193452
G_{pp} (eV)		17.042124	14.582090	13.641224
G_{p2} (eV)		10.338310	10.059431	12.571464
H_{sp} (eV)		2.811269	3.616781	0.589834
α (Å^{-1})	2.739361	5.799824	2.278671	1.498274
K_1 (eV)	0.049143	0.352574	−0.354839	
L_1 (Å^{-1})	3.953000	3.030000	6.051865	
M_1 (Å)	0.980400	0.974529	1.255455	
K_2 (eV)	0.003321	0.048729	−0.013484	
L_2 (Å^{-1})	8.256000	6.150900	5.360431	
M_2 (Å)	1.345680	1.841512	0.695556	
K_3 (eV)	−0.021350		0.015632	
L_3 (Å^{-1})	1.897200		11.150605	
M_3 (Å)	1.270080		1.240743	

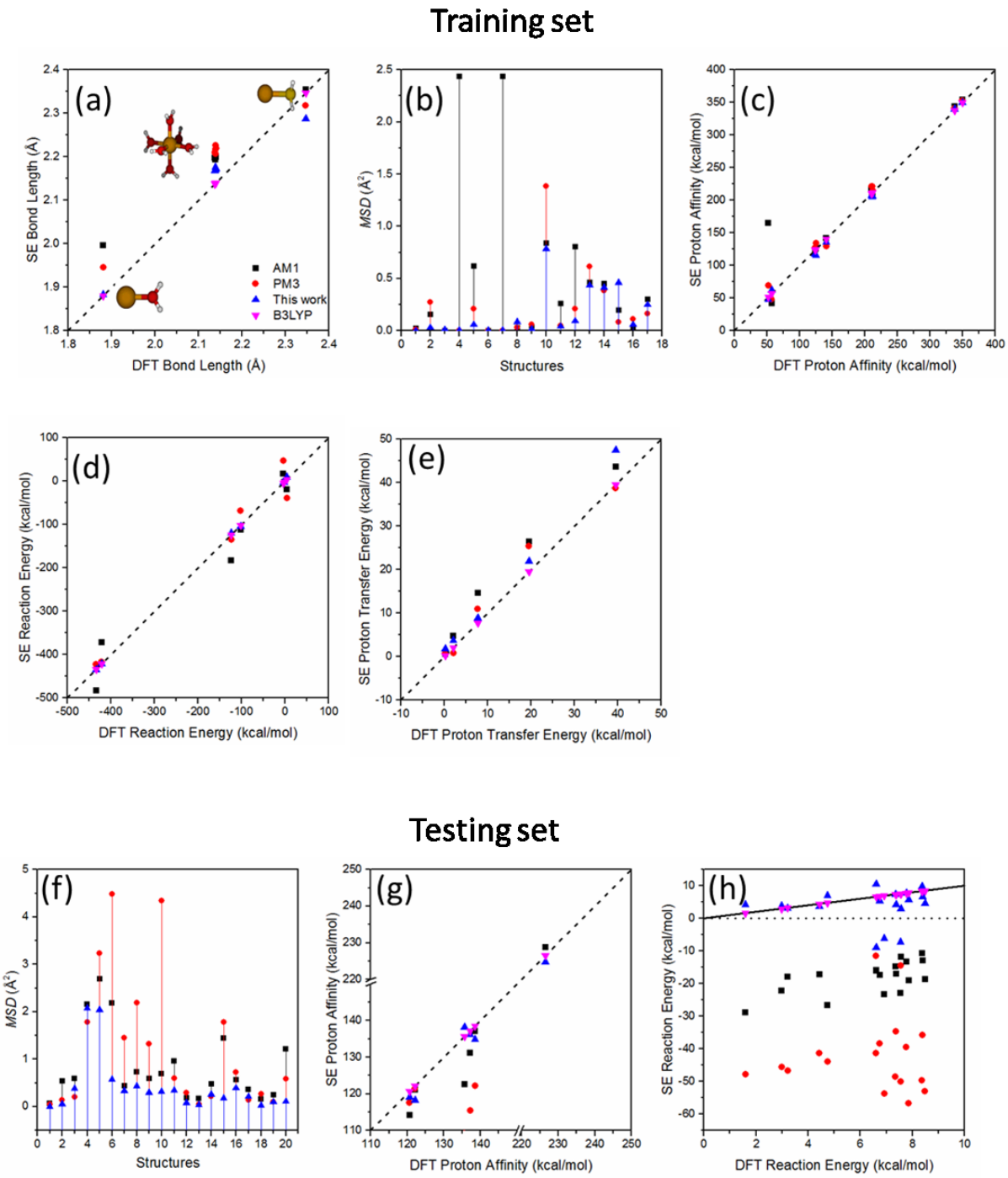


Figure 4.3: Performance of AM1, PM3, and the optimized models on small clusters of the training and testing sets. The dotted diagonal and inclined solid lines inside the figures represents the correlation of SE results with respected to DFT. Black, red, blue and pink symbols denote for AM1, PM3, This work and B3LYP results.

Figure 4.3 (f-h) presents performance of the optimized models at reproducing structures, proton affinities, and reaction energies in comparison with the performance of AM1 and PM3. As shown in Figure 4.3 (f-h) and Table 4.2, the optimized model gives low mean square deviations in cluster geometries in reference to B3LYP optimized geometries. Compared to AM1 and PM3, the optimized model yields proton affinities and reaction energies in better agreement with B3LYP results. The optimized model yields average errors for MSDs, proton affinities and ligand exchange reactions of 0.41 $Å^2$, 2.42 kcal/mol and 4.02 kcal/mol respectively. In comparison, the original AM1 model shows errors of 0.82 $Å^2$, 5.08 kcal/mol, and 24.29 kcal/mol and the PM3 model gives average errors of 1.20 $Å^2$, 13.00 kcal/mol, and 48.19 kcal/mol. The optimized model yields three incorrect ligand exchange energies, which are associated with the facial and meridional configuration of $Zn^{2+}{\cdot}3H_2O{\cdot}3H_2S$ going to cis $Zn^{2+}{\cdot}2H_2O{\cdot}4H_2S$. The large errors in energies are due to the distorted structure of cis-$Zn^{2+}{\cdot}2H_2O{\cdot}4H_2S$, where one H_2S leaves the first coordination shell and forms new hydrogen-bond with another H_2S. Other large error is coming from the ligand exchange from trans- $Zn^{2+}{\cdot}2H_2O{\cdot}4H_2S$ going to $Zn^{2+}{\cdot}1H_2O{\cdot}5H_2S$. As previously, large error in energy is coming from the severe distortion of octahedral structure of $Zn^{2+}{\cdot}1H_2O{\cdot}5H_2S$, where one H_2S leaves first coordination shell and forms a hydrogen bond with another H_2S. From a biological standpoint, these hexacoordinated zinc clusters with four or five sulfur-containing ligands have limited significance.[200] For both AM1 and PM3, the large errors are due to distorted structure. Moreover, PM3 (1.29 Å) gives shorter H–S than B3LYP (1.35 Å) in H_2S. The large errors in proton affinities in zinc-bound water complexes from PM3 may arise from incorrect ∠Zn···O–M angle (M is the point on H–O–H angle bisector), which is 31° smaller than in B3LYP structures.

Table 4.2: Average unsigned errors for each property used in the training and testing sets.

	Properties	*N**	Average unsigned errors		
			AM1	PM3	This work
Training set	MSDs (Å^2)	17	0.53	0.21	0.16
	Bonds (Å)	8	0.06	0.07	0.03
	Proton affinities (kcal/mol)	10	15.13	7.86	4.53
	Proton transfer energies (kcal/mol)	5	4.21	2.21	2.92
	Reaction energies (kcal/mol)	7	30.30	22.30	2.27
Testing set	MSDs (Å^2)	20	0.82	1.20	0.41
	Proton affinities (kcal/mol)	6	5.08	13.00	2.42
	Reaction energies (kcal/mol)	18	24.29	48.19	4.02

*N represents the number of terms used in the error function

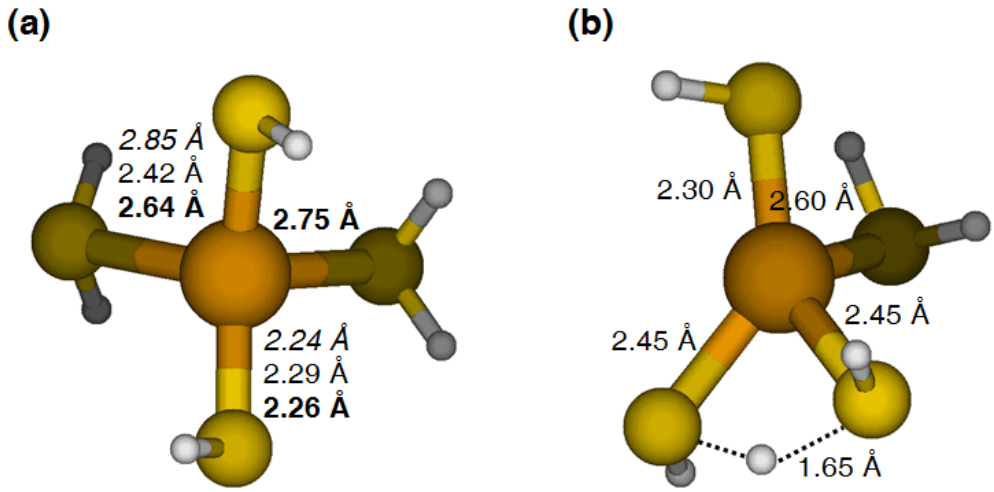


Figure 4.4: Zn···S^0/S^{-1} distances of the $Zn^{2+}\cdot 2H_2S\cdot 2HS^-$ complex as obtained from B3LYP (a, italic), PM3 (a, normal), the optimized model (a, bold; two different Zn···S^0 distances of 2.64 and 2.75 Å are found), and the original AM1 model (b, normal). The latter model predicts a wrong geometry of the complex.

In $Zn^{2+}\cdot 4H_2S$, all models predict tetrahedral geometry with Zn···S distances of 2.52 (AM1), 2.39 (PM3), and 2.47 Å (the optimized model) in good agreement with 2.46 Å for B3LYP. Two different bond distances ($Zn\cdots S^0$ and $Zn\cdots S^-$) are found in the $Zn^{2+}\cdot 2H_2S\cdot 2HS^-$ complex (see Figure 4.4). As presented in Figure 4.4, PM3 significantly underestimates $Zn\cdots S^0$ (2.42 Å) distance and slightly overestimates $Zn\cdots S^-$ (2.29 Å) as compared to the B3LYP values of 2.85 and 2.24 Å respectively. However, AM1 produces a distorted tetra-coordinated structure, where one hydrogen atom is shared by two sulfur atoms. The optimized model yields $Zn\cdots S^0$ (2.64 and 2.75 Å) and $Zn\cdots S^-$ (2.26 Å) bond distances in good agreement with the B3LYP values.

In the case of $Zn^{2+}\cdot 4HS^-$, both AM1 and the optimized model yield comparable four $Zn\cdots S^-$ bond distances of 2.46 and 2.44 Å, respectively, consistent with B3LYP value of 2.45 Å. However, PM3 shows a smaller value of 2.39 Å. Two different values of S···Zn···S angle are observed with values 106° and 116° in B3LYP optimized structure. Compared with DFT values, AM1 underestimates one with 104° and overestimates the other with 121°. Despite underestimation of bond lengths, PM3 reproduces both angles of values 108° and 112° consistent with DFT results. The optimized model reproduces both angles (108° and 111°) in good agreement with B3LYP values. This suggests that the optimized model yields gas-phase structures of zinc-H_2S complexes in their different charged states in good agreement with B3LYP results.

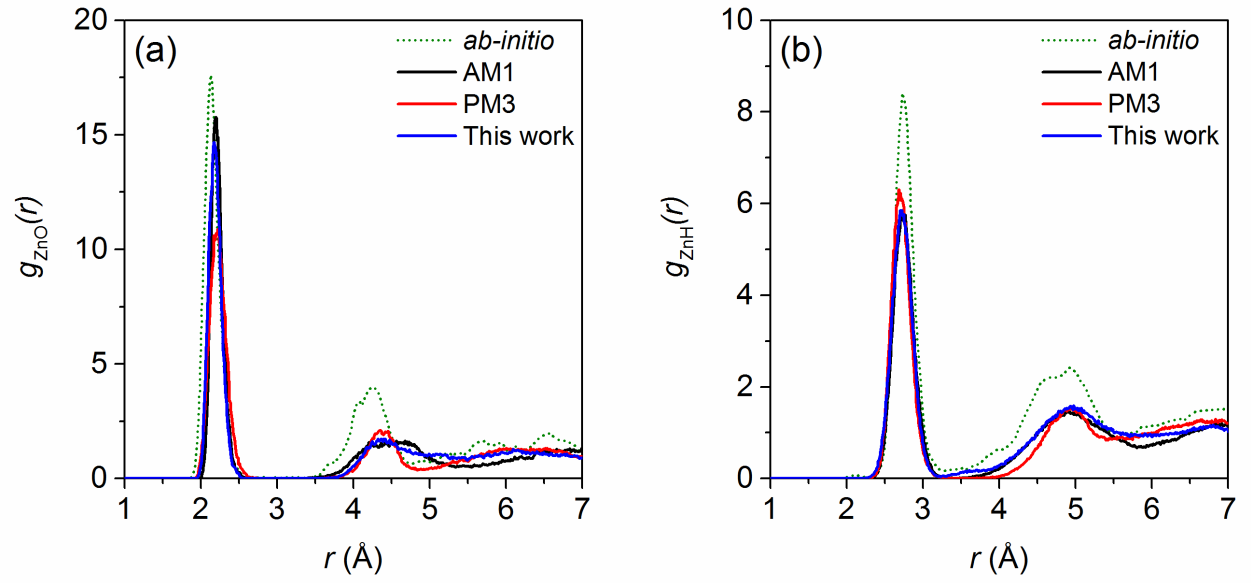


Figure 4.5: Radial distribution functions of (a) Zn–O and (b) Zn–H pairs for zinc solvated in 63 water molecules from AM1 (black), PM3 (red), and the optimized (blue) model in comparison with ab initio (dotted green) results[224].

4.4.2 Hydration structure of Zn^{2+}

Consistent with experiment[225] and ab initio results[224], all three SE models predict an hexa-coordinated hydration structure for Zn^{2+}. Figure 4.5 shows the calculated Zn–O and Zn–H radial distribution functions (RDFs), $g_{\mathrm{ZnO}}(r)$ and $g_{\mathrm{ZnH}}(r)$. As shown in Figure 4.5 (a) and Table 4.3, the three SE models give $g_{\mathrm{ZnO}}(r)$ that displays a well-separated, sharp first peak centered at ~2.2 Å, in good agreement with experimental (2.06 Å)[225] and ab initio (2.12 Å)[224] values. Relative to ab initio results, the PM3 model underestimates the intensity of the first peak of $g_{\mathrm{ZnO}}(r)$. The trend in the intensity of the first peak (AM1 > optimized model > PM3) is consistent with the trend in binding affinity predicted by the models for the Zn^{2+}-H_2O dimer (–112.4, –104.1, –67.8 kcal/mol). The function displays a second peak that, compared to the first peak, is broad and less intense. All three models yield well-defined second solvation shell of zinc ion. It is somehow surprising that AM1 also reproduces position of the second peak, given that it poorly describes hydrogen bonding between water molecules. The peaks in the $g_{\mathrm{ZnO}}(r)$ are well separated and in fact, we have not observed any exchange for the ion's first shell water molecules within the 40-ps

simulation time. The three models yield similar $g_{\mathrm{ZnH}}(r)$ (see Figure 4.5 (b) and Table 4.3) and reproduce well-defined first and second solvation shells similar to the ab initio function.

Table 4.3: Characteristic values of radial distribution functions obtained from *NVT* simulations of zinc ion solvated with 63 water molecules using AM1, PM3, and the optimized model in comparison with theoretical and experimental results. r^{max1} and r^{max2} are the locations of first and second maxima on each curve. N_{O} is the oxygen coordination number, integrated from r_{ZnO} equal to zero to the first minimum on each curve.

Methods	Zn–O			Zn–H	
	$r_{\mathrm{ZnO}}^{\mathrm{max1}}$ [Å]	$r_{\mathrm{ZnO}}^{\mathrm{max2}}$ [Å]	N_{O}	$r_{\mathrm{ZnH}}^{\mathrm{max1}}$ [Å]	$r_{\mathrm{ZnH}}^{\mathrm{max2}}$ [Å]
Ab initio MD[a]	2.12	4.24	6	2.75	4.92
AM1	2.20	4.50	6	2.77	4.92
PM3	2.23	4.34	6	2.66	4.95
This work	2.17	4.32	6	2.73	4.95
Expt.[b]	2.06		6		
AIMD and AIMD/MM[c]	2.18, 2.19	4.22, 4.31	6	2.75, 2.77	4.62,4.71

(a) Ref. [224] (b) Ref. [225] (c) Ref. [226]

4.4.3 Zn^{2+} in liquid hydrogen sulfide

AM1 and PM3 do not consistently reproduce gas-phase zinc-sulfur bond lengths[131] (also see Figure 4.4). To explore zinc-sulfur coordination in condensed phase, the overall solvation structure of zinc ion in liquid hydrogen sulfide is studied from AM1, PM3, and the optimized model. The radial distribution functions, $g_{\mathrm{ZnS}}(r)$ and $g_{\mathrm{ZnH}}(r)$, along with respective running coordination numbers are analyzed.

As shown in Figure 4.6 (a), PM3 and the optimized model show a clearly defined first solvation shell. AM1 however, displays no clear separation between the first and second

solvation shells. B3LYP calculations suggest that ~30 kcal/mol is liberated when zinc–hydrogen sulfide complex goes from tri-coordination to tetra-coordination and that zinc does not form a stable penta-coordinated complex. This suggests that zinc prefers to bind four hydrogen sulfide molecules instead of three or five. In proteins, 40% of all tetrahedrally coordinated zinc atoms bind four cysteines.[227] A survey on metal binding protein coordination also shows that zinc binds only three or four cysteines.[45]Thus, it is expected, from the MD simulations, that zinc binds four hydrogen sulfide molecules in the first coordination shell. However, both AM1 and PM3 favor zinc binds five hydrogen sulfide molecules in the first solvation shell (see Table 4.4). Unlike AM1 and PM3, the optimized model yields a well-defined first solvation shell in which zinc binds four hydrogen sulfides. Unlike PM3 and the optimized model, AM1 also shows early second solvation shell around 3.70 Å showing strong zinc-sulfur interactions. As shown in Figure 4.6 (b), AM1 and the optimized model show clear first maxima and minima of $g_{\mathrm{ZnH}}(r)$ but not PM3.

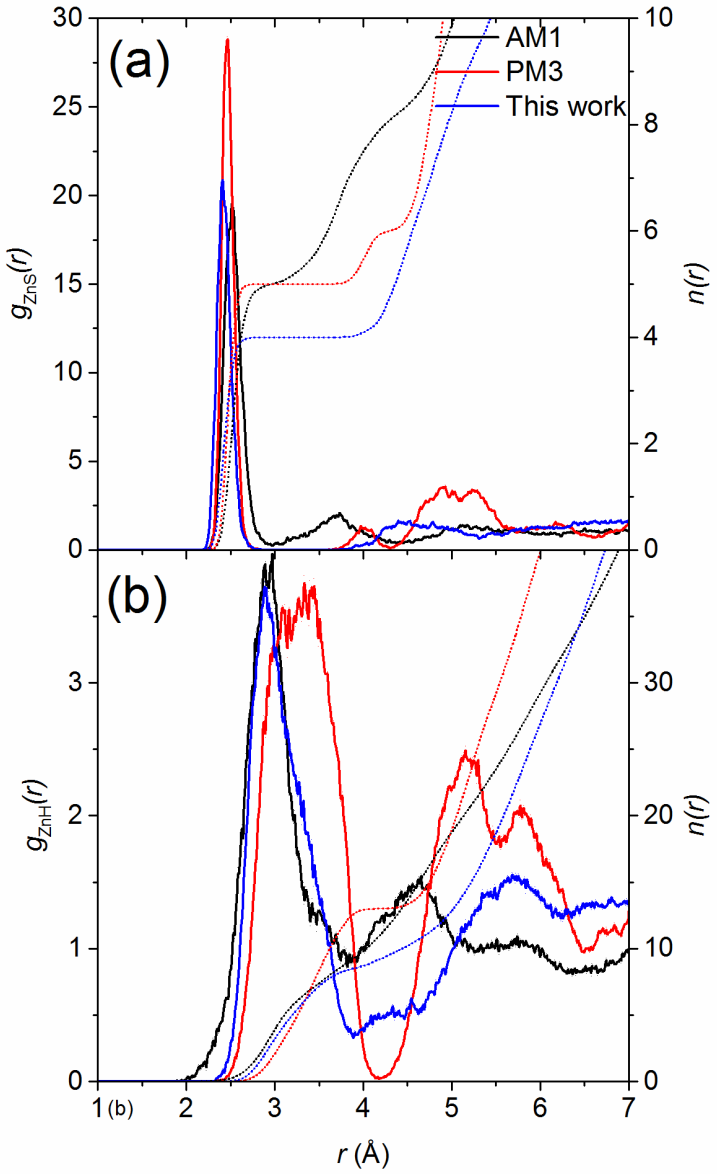


Figure 4.6: Radial distribution functions, $g(r)$, of Zn–S (panel a) and Zn–H (panel b) for zinc solvated in 63 hydrogen sulfide molecules at T = 300 K from AM1 (black), PM3 (red), and the optimized model (blue). Dotted lines represent $n(r)$, the corresponding running coordination numbers.

Table 4.4: Characteristic values of radial distribution functions obtained from *NVT* simulations of solvation of zinc ion in liquid hydrogen sulfide of 63 molecules using AM1, PM3, and the optimized model. Symbols have their usual meanings.

Methods	Zn–S			Zn–H	
	$r_{\mathrm{ZnS}}^{\mathrm{max1}}$ [Å]	$r_{\mathrm{ZnS}}^{\mathrm{max2}}$ [Å]	N_{S}	$r_{\mathrm{ZnH}}^{\mathrm{max1}}$ [Å]	$r_{\mathrm{ZnH}}^{\mathrm{max2}}$ [Å]
AM1	2.51	3.70	5.0	2.93	4.62
PM3	2.46	4.02	5.0	3.29	5.18
This work	2.42	4.64	4.0	2.89	5.68

4.4.4 Zinc-bound H_2S/HS^- in water

MD simulations using AM1, PM3, and the optimized model were performed on zinc-bound H_2S and zinc-bound HS^- solvated in 62 water molecules. Figure 4.7 presents the radial distribution functions, $g_{\mathrm{ZnO}}(r)$, and running integration numbers, $n(r)$, of zinc-oxygen pairs. Average Zn···S^0 and Zn···S^- distances in these two systems are also analyzed from the radial distribution functions, $g_{\mathrm{ZnS}}(r)$, as shown in Figure 4.8.

As shown in Figure 4.7 (a) and Table 4.5, position of the first peak of RDFs of Zn···O pair from AM1, PM3 and the optimized model are 2.20, 2.09, and 2.18 Å respectively. In the case of AM1 and the optimized model, the presence of H_2S in the first coordination shell does not change position of the first peak of RDF of Zn···O pair significantly. However, this is not the case for PM3, where the first peak position is significantly shortened to 2.09 Å as compared to 2.23 Å in neat water. The binding energies from B3LYP suggest that hexa-coordinated $Zn^{2+}{\cdot}5H_2O{\cdot}H_2S$ (–340.7 kcal/mol) is 22.1 kcal/mol more stable than penta-coordinated $Zn^{2+}{\cdot}4H_2O{\cdot}H_2S$ (–318.6 kcal/mol) complex. Thus, it is expected that zinc ion retains the octahedral coordination at the first shell when bound to neutral H_2S. The number of oxygen (N_{O}) in the first shell indicates that AM1 (4.0) and PM3 (3.5) do not maintain the octahedral first solvation shell. However, the optimized model retains hexacoordinated structure with five water

molecules and one hydrogen sulfide. No water and hydrogen sulfide exchange were observed throughout the simulation period.

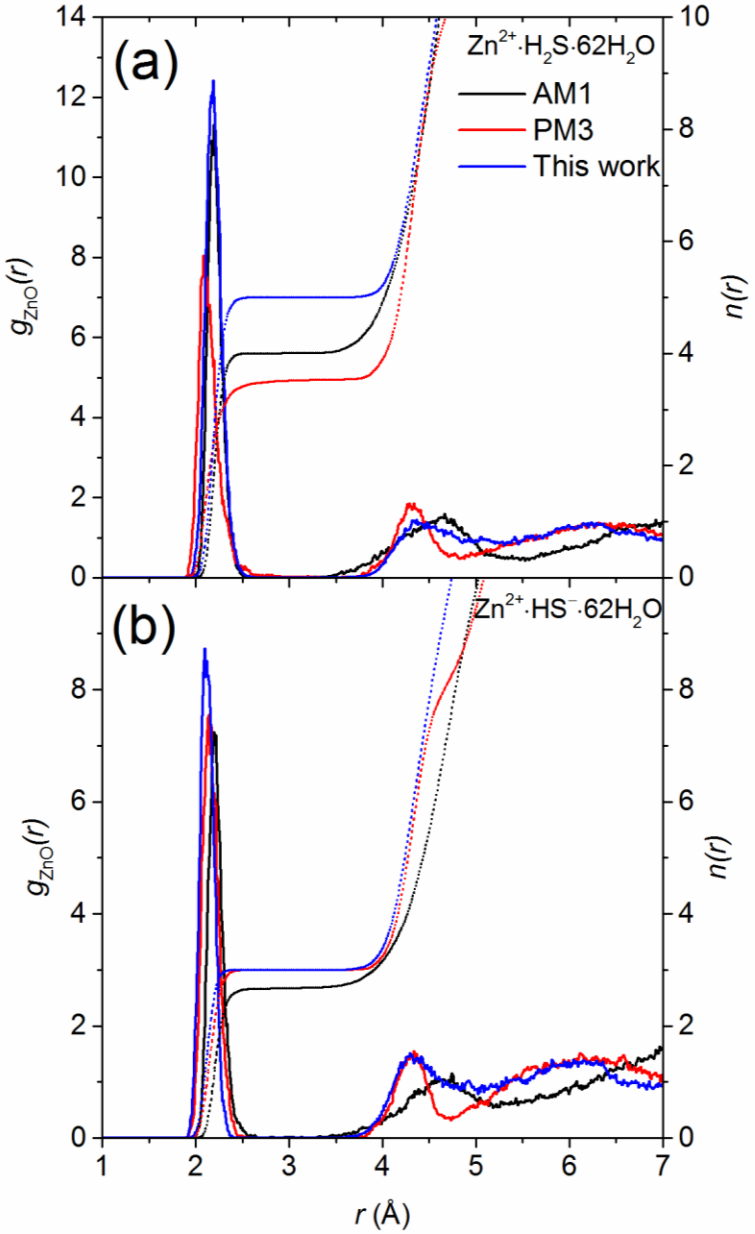


Figure 4.7: Radial distribution functions, $g(r)$ of Zn with O obtained from AM1 (black), PM3 (red), and the optimized model (blue) solvated in one hydrogen sulfide and 62 water molecules at T=300 K. Upper and lower panels show the Zn···O RDFs in zinc solvated in water when zinc binds neutral and deprotonated hydrogen sulfide. Dotted lines represent the corresponding running coordination numbers.

Table 4.5: Results of *NVT* simulations of zinc-bound H_2S and zinc-bound HS^- in liquid water of 62 molecules obtained from AM1, PM3, and the optimized model.

Methods	$Zn^{2+}\cdot H_2S\cdot 62H_2O$			$Zn^{2+}\cdot HS^-\cdot 62H_2O$		
	r_{ZnO}^{max1} [Å]	r_{ZnS}^{max1} [Å]	N_O	r_{ZnO}^{max1} [Å]	r_{ZnS}^{max1} [Å]	N_O
AM1	2.20	2.41	4.0	2.20	2.20	2.7
PM3	2.09	2.37	3.5	2.15	2.24	3.0
This work	2.18	2.64	5.0	2.09	2.25	3.0

For $Zn^{2+}\cdot HS^-\cdot 62H_2O$ system, as shown in Figure 4.7 (b) and Table 4.5, average Zn···O distance (2.20 Å) remains unchanged for AM1 and decreases slightly to 2.15 Å for PM3. For both AM1 and PM3, the water coordination numbers are decreased to 2.7 and 3.0 respectively. Metal-oxygen interactions are expected to increase as the inner shell coordination changes from octahedral to tetrahedral, resulting in a shortening of zinc-oxygen bonds. This behavior is not observed for AM1 and PM3. Zinc cation interacts stronger with the deprotonated ligand than with the neutral. For instance, the binding energies from B3LYP suggests that $Zn^{2+}\cdot 3H_2O\cdot HS^-$ (–522.9 kcal/mol) is 228.6 kcal/mol more stable than $Zn^{2+}\cdot 3H_2O\cdot H_2S$ (–294.3 kcal/mol). Unlike AM1 and PM3, the optimized model systematically reproduces the coordination geometry of the first shell. For example, zinc binds three water molecules in the first coordination shell and forms a tetrahedral structure along with HS^-. The Zn···O (2.09 Å) distance, in this case, is shorter than the Zn···O (2.18 Å) distance when zinc binds neutral hydrogen sulfide.

Figure 4.8 presents $g_{ZnS}(r)$, the radial distribution functions of zinc-sulfur pairs in $Zn^{2+}\cdot H_2S\cdot 62H_2O$ and $Zn^{2+}\cdot HS^-\cdot 62H_2O$ systems. From Table 4.5, it is found that all models produce shorter Zn–S^- than Zn–S^0 due to stronger electrostatic interaction between zinc and sulfur in $Zn^{2+}\cdot HS^-\cdot 62H_2O$ than in $Zn^{2+}\cdot H_2S\cdot 62H_2O$. Compared to other models, the optimized model shows a broader distribution of Zn–S^0 distances from ~2.35 Å to ~3.00 Å. As presented in Table 4.5, AM1 and PM3 yield Zn–S^0 and Zn–S^- distances (r_{ZnS}^{max1}) of 2.41 and 2.20 Å and 2.37

and 2.24 Å respectively. The optimized model, however, yields both Zn–S^0 (2.64 Å) and Zn–S^- (2.25 Å) distances consistent with gas-phase B3LYP results.

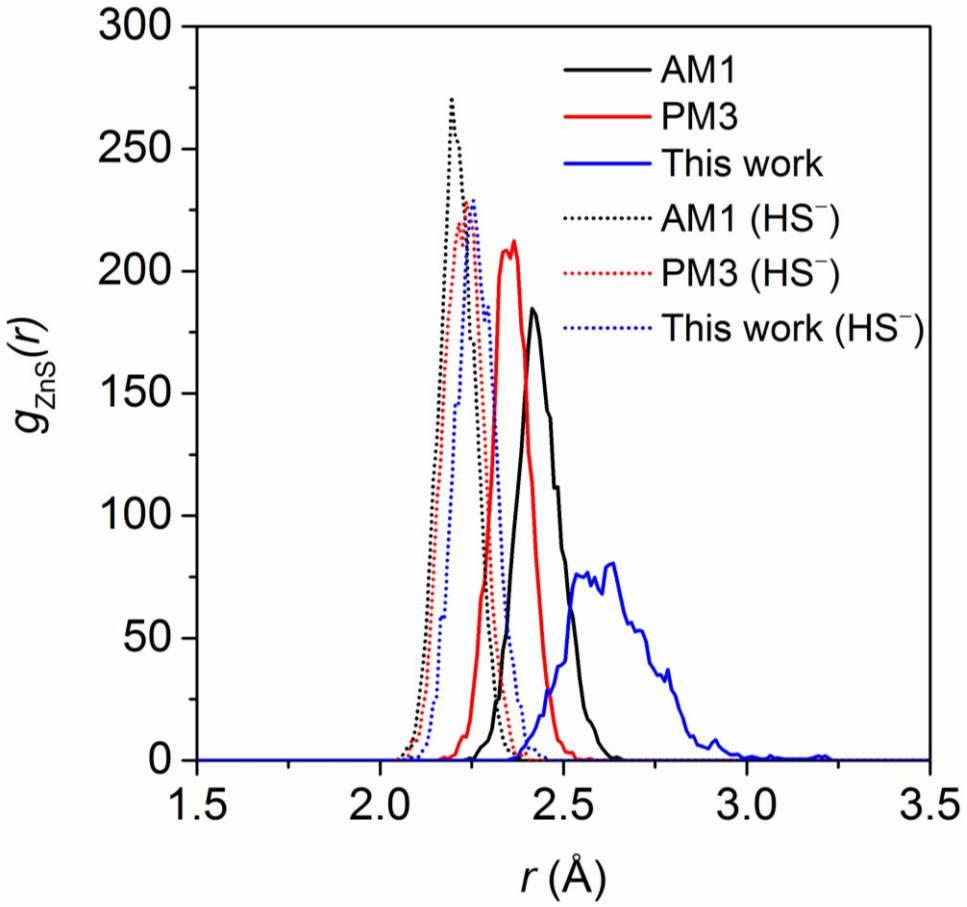


Figure 4.8: Radial distribution function, *g(r)*, of Zn-S (solid line) and Zn–S^- (dotted line) obtained from MD simulations of $Zn^{2+}\cdot H_2S\cdot 62H_2O$ and $Zn^{2+}\cdot HS^-\cdot 62H_2O$ respectively. Results from AM1 (black), PM3 (red), and the optimized model (blue) are presented for both systems.

4.4.5 Angular distribution functions

Angular distribution functions provide an easy-to-interpret signature of the coordination geometry of the ion. To understand what local structure is depicted by a zinc ion, Zn atom centered angular distribution functions, ∠OZnO and ∠OZnS (for all O and S atoms in the first shell) are analyzed. Maxima around 90° and 180° indicate an octahedral arrangement of the solvent molecules whereas a broadened peak indicates distorted basic octahedral structure.

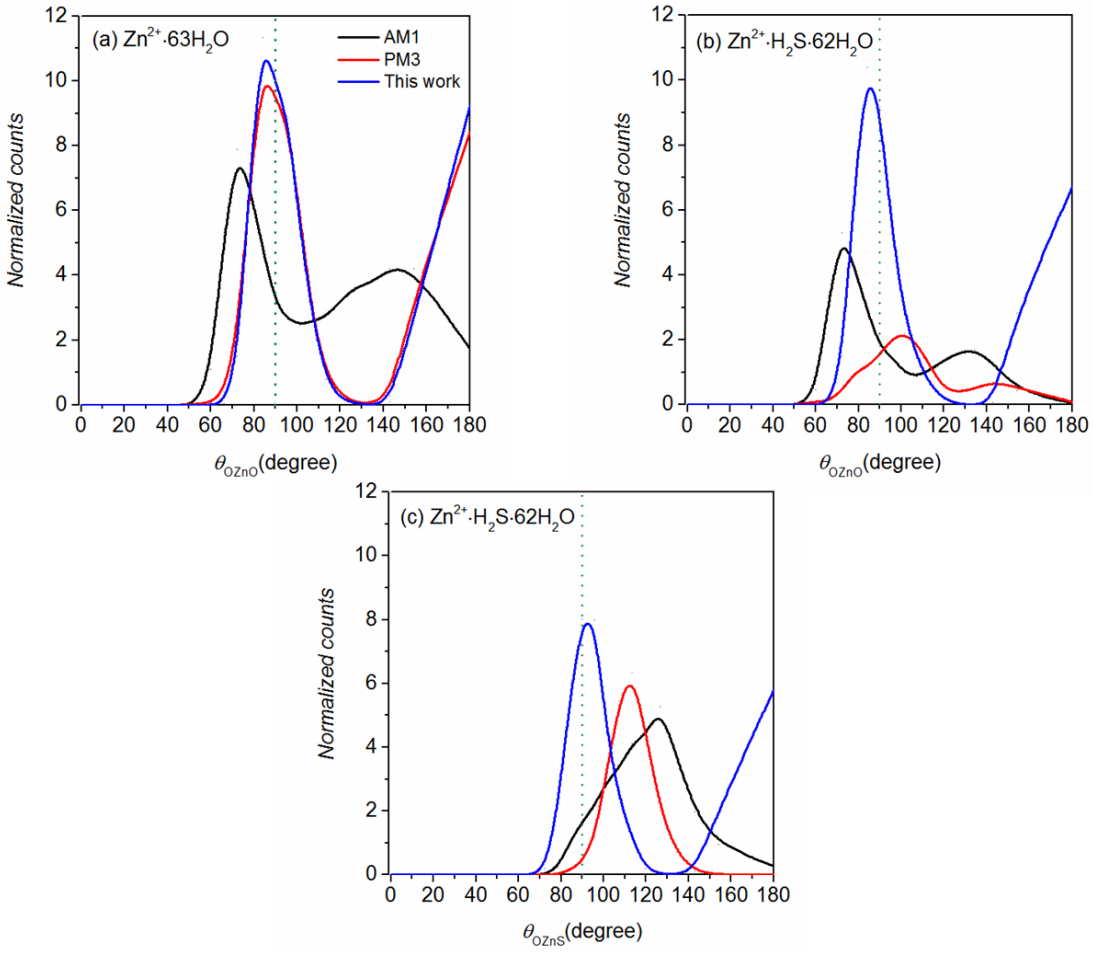


Figure 4.9: Angle distribution functions for ligands in the first solvation shell of zinc obtained from AM1 (back), PM3 (red) and the optimized model (blue) within Zn···O/S≤3.0 Å in systems (a) for θ_{OZnO} in $Zn^{2+}\cdot 63H_2O$, (b) for θ_{OZnO} in $Zn^{2+}\cdot H_2S\cdot 63H_2O$, and (c) for θ_{OZnS} in $Zn^{2+}\cdot H_2S\cdot 62H_2O$. The vertical green line represents θ=90°. Peaks at θ=90° and 180° represent octahedral structure.

Figure 4.9 presents the O···Zn···O angle distribution functions in $Zn^{2+}\cdot 63H_2O$ and $Zn^{2+}\cdot H_2S\cdot 62H_2O$ and O···Zn···S angle distribution functions in $Zn^{2+}\cdot H_2S\cdot 62H_2O$ computed from AM1, PM3, and the optimized model. All distribution functions were computed from the first solvation shell defined as Zn···O/S ≤ 3.0 Å.

Both PM3 and the optimized model predict an octahedral hydration structure of Zn^{2+} (Figure 4.9 (a)). In comparison, the original AM1 model predicts a distorted octahedral structure, suggesting an overly flexible coordination. Panels b and c of Figure 4.9 show that, while AM1 and PM3 predict that a single H_2S molecule in the first solvation shell of Zn totally distorts the octahedral solvation structure, the optimized model retains the octahedral structure of the ion in the presence of a single H_2S molecule.

4.4.6 Metal-induced pK_a shifts

Calculating accurate pK_a values from free energy profiles is very challenging because of the fact that small errors in free energies can lead to large errors in pK_a values.[228] We are calculating pK_a from the method of coordination constraints dynamics,[229] which is one of the numerically reliable methods for the hydrogen bonded systems of short-time-scale dynamics. Metal-induced pK_a shifts of water were computed from the free energy profiles of deprotonation of a zinc-bound water.[101] The deprotonation free energy profile was computed by imposing a constraint on a reaction coordinate that drives the transfer of a proton from a zinc-bound water to a second-shell water molecule forming a hydrogen bond with it. Molecular dynamic simulations were performed by imposing distance constraints of r_{OH} (between hydrogen of proton-donating water molecule in first shell and oxygen of proton-accepting water molecule in second shell), from 1.0 to 1.6 Å with a step size of 0.1 Å. For each constraint, average Lagrange multiplier was obtained from five independent 40-ps long MD simulations (10-ps equilibration and 30-ps production). For the pure water system, average Lagrange multipliers are calculated from two independent MD simulations. The mean force and the free energy profiles were then calculated. See ref. [101] and section 2.7 for details of pK_a calculation. To calculate pK_a of H_2S in water and zinc-induced pK_a shift of H_2S in water, a 40-ps long MD simulations (10-ps equilibration and 30-ps production) was used to calculate average Lagrange multiplier for each constraint in each system. This was carried out by imposing constraints on the distance between H from H_2S and O from a neighboring hydrogen-bonded water molecule from 1.0 to 2.2 Å in increments of 0.2 Å.

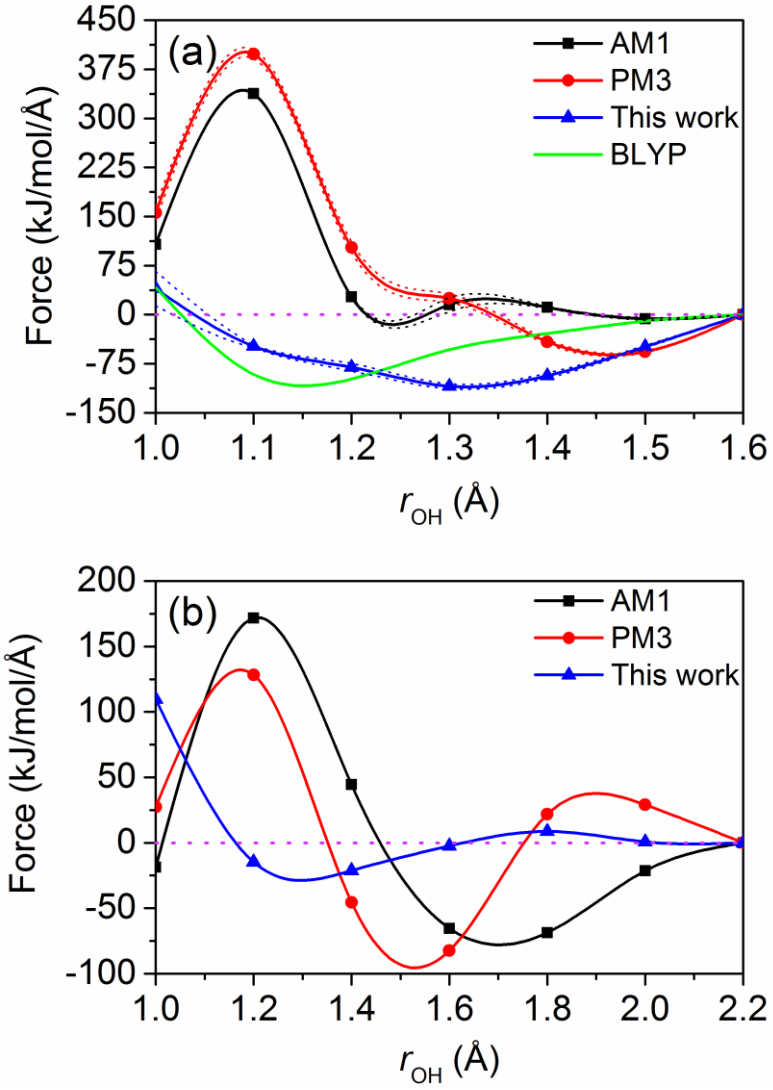


Figure 4.10: (a) Mean force of the deprotonation of zinc-bound water along the imposed mechanical constraints of r_{OH} from 1.0 to 1.6 Å. BLYP results were adapted from ref. [101] for comparison. (b) Mean force of the deprotonation of zinc-bound H_2S in liquid water by constraining between H from H_2S and O (from the water molecule forming a hydrogen bond with the H_2S bound to zinc) from 1.0 to 2.2 Å. Presented mean forces are relative to the forces at r_{OH}= 1.6 Å (panel a) and r_{OH} = 2.2 Å (panel b). Dotted lines (panel a) are the estimated standard errors obtained from multiple independent simulations. Lines are obtained from spline interpolation of the calculated values (shown with symbols). The horizontal purple dotted lines are the zero forces.

Figure 4.10 presents the relative mean forces as a function of order parameter, r_{OH}, obtained from AM1, PM3, and the optimized model. As shown in Figure 4.10 (a), both AM1 and PM3 yield unrealistic mean force profiles of deprotonation of zinc-bound water in water. The optimized model, however, slightly overestimates the position of minima but yields the acceptable overall nature of the profile in comparison with BLYP. Similarly, both AM1 and PM3 fail at reproducing reasonable mean force profile in the case of zinc-bound hydrogen sulfide as shown in Figure 4.10 (b). The failure of predicting a correct proton transfer energies profiles from AM1 and PM3 may arise from the inability of reproducing correct gas-phase proton affinities of zinc bound ligands. For instance, the gas-phase proton affinities of zinc-bound H_2O and H_2S from AM1 (42.30 and 165.45 kcal/mol) and PM3 (48.20 and 69.62 kcal/mol) are not consistent with B3LYP (57.07 and 51.78 kcal/mol) results. The force profiles for AM1 and PM3 are qualitatively incorrect, and will not be analyzed further. Unlike AM1 and PM3, the optimized model gives proton affinities of zinc-bound H_2O (62.01 kcal/mol) and H_2S (48.06 kcal/mol) in good agreement with B3LYP results and yields acceptable mean force profiles in liquid.

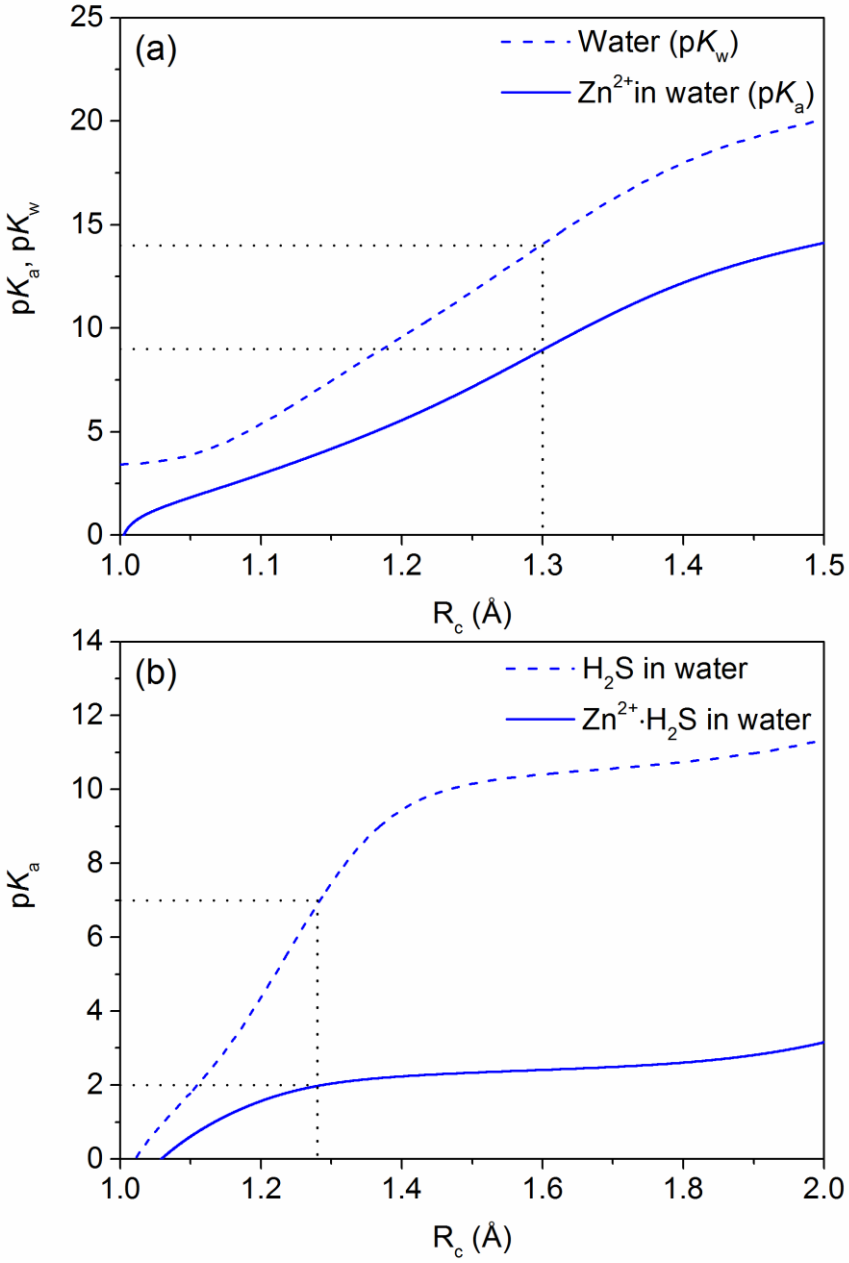


Figure 4.11: Estimated pK_a values as a function of the bonding radius R_c to find pK_a shift of (a) zinc-bound water and (b) zinc-bound hydrogen sulfide. Vertical dotted lines show R_c and horizontal dotted lines shows corresponding pK_a values.

The estimated pK_a values as a function of the bonding cutoff radius R_c (R_c is the order parameter at which the bond with the donor molecule is considered broken) for zinc-bound water,

zinc-bound hydrogen sulfide, and hydrogen sulfide in water calculated using eq. (2-53) and eq. (2-56) and pK_w for the ionic product of water is computed using eq. (2-55) and (2-56) are presented in Figure 4.11. Optimal R_c values are estimated by assuming that the model reproduces the correct experimental pK_a of clean water and hydrogen sulfide in water.

R_c is chosen based on the proton transfer in neat water (Figure 4.11 (a)), such that pK_w(R_c) = 14. This cutoff radius is then used to estimate the pK_a of zinc-bound water. The model yields a cutoff radius R_c = 1.30 Å in good agreement with the previously reported value of 1.28 Å[101] from ab initio molecular dynamics (AIMD) simulations at the BLYP level of theory. Furthermore, with the value of R_c = 1.30 Å, the optimized model estimates a pK_a value of 9.0 for zinc-bound water. This estimate is in very good agreement with the experimental values from the literature: 8.96 (at 25°C),[230] 9.0 (at 25°C),[231] 9.5,[232] and 9.6.[233] The reduction of pK_a of water upon zinc coordination (from 14 to 9.0) is due to the electrostatic stabilization of hydroxide ion by Zn^{2+} cation. In a biological environment, specifically in zinc metalloproteins, however, the pK_a value can further shift down due to lower coordination environment (hexa to tetra) of zinc, charges of amino acid side chains, and overall dielectric environment at the active site. Due to the smaller ionic radius of zinc in a tetrahedral configuration (0.74 Å)[233] than in an octahedral configuration (0.88 Å)[233], pK_a of zinc-bound water can be shifted down by additional 2.5 pK_a units so that water can spontaneously ionize at physiological pH.[233]

In aqueous solution, hydrogen sulfide dissociates in two steps: from neutral (H_2S) to hydrosulfide anion (HS^-), and from hydrosulfide anion (HS^-) to sulfide anion (S^{2-}). In aqueous solution, the pK_a for the first dissociation reaction ($H_2S \rightleftharpoons H^+ + HS^-$) is close to physiological pH (6.95 (at 25°C),[234] 6.98 (at 25°C),[235] 7.04,[236,237] and 7.05 (at 25°C)[238]) whereas pK_a for the second dissociation ($HS^- \rightleftharpoons H^+ + S^{2-}$) is very high (11.96[237] and 19±2[235]). At physiological pH, more than two-thirds of hydrogen sulfide form hydrosulfide anion, whereas the remaining third is in neutral form. Second-step dissociation only occurs at high pH and it does not exist *in vivo*[239] hence it has no biological significance. In this work, pK_a is estimated for the first dissociation reaction of zinc-bound hydrogen sulfide in water. Figure 4.11 (b) presents the pK_a profiles for hydrogen sulfide and zinc-bound hydrogen sulfide in water. Similarly to zinc-bound water, it is assumed that the optimized model reproduces the experimental pK_a of H_2S in water. Average experimental pK_a value of 7.0 is used to estimate R_c of H_2S in water. As shown in Figure 4.11 (b),

the estimated R_c value is 1.28Å which is acceptable proton dissociation distance based on the results for zinc-bound water. Based on this R_c value, the model estimates a pK_a of 2.0. The optimized model yields pK_a of H_2S, which is similar to pK_a of cysteines (1.9 and 2) in protein environment. [240] Both H_2S and cysteine have been shown to have comparable proton affinities and H_2S has been used to model the cysteine side chain.[241] Moreover, the optimized model predicts that zinc cation lowers the pK_a of hydrogen sulfide upon bound to it in water by 5 pK_a units. This result is consistent with previous findings in proteins, where zinc lowers the pK_a of cysteines by 4-7 units.[240]

4.5 Conclusion

AM1 parameters for Zn and S were optimized, using a genetic algorithm approach, to reproduce geometries, ligand-exchange energies, proton transfer energies, and proton affinities of simple zinc-bound water and hydrogen sulfide complexes mimicking the active sites of zinc metalloproteins. Reference data were obtained from hybrid density functional method B3LYP. Previously optimized parameters for H and O to describe proton transfer reactions in water were used. It is found that the optimized model performs better than original AM1 and PM3 models for both training and testing sets. In particular, AM1 gives unrealistic distorted $Zn^{2+}\cdot 2H_2S\cdot 2HS^-$ complex in which one hydrogen atom is shared by two neighboring sulfur atoms. Compared to B3LYP results, PM3 significantly underestimates the zinc-sulfur bond distances. From molecular dynamics simulations, it is found that the optimized model produces a hydration structure of zinc ion in line with high-level theoretical and experimental results. While yielding correct coordination number for hydrated zinc ion, AM1 and PM3 do not reproduce the location of the first peak of the zinc-oxygen radial distribution function. Unlike AM1 and PM3, the optimized model also reproduces a coordination of zinc in liquid hydrogen sulfide consistent with the coordination geometry found at the structural active sites of zinc metalloproteins.

The hydration structures of zinc ion with and without hydrogen sulfide were analyzed. It is found that AM1 and PM3 do not reproduce the octahedral coordination of the ligands in first coordination shell of zinc ion. Both models show zinc ion loses its hexa-coordination upon introduction of hydrogen sulfide in the first shell. In contrast, the optimized model consistently retains the local octahedral symmetry.

Furthermore, zinc-induced pK_a shift of water was investigated. The optimized model estimates the pK_a shift of zinc-bound water in excellent agreement with experimental and theoretical values. The findings suggest that the model can accurately describe the acidity of zinc-bound water at active sites of zinc metalloproteins. The same procedure was used to calculate pK_a shift of zinc-bound hydrogen sulfide in water and the pK_a value obtained is consistent with pK_a of catalytic cysteine in protein environment. This suggests that hydrogen sulfide can be used as a model compound for cysteine.[241] The optimized model provides a significant improvement over original AM1 and PM3 models to describe various properties in gas-phase and liquid-phase. This work also confirms that the used parameterization technique is reliable to design new semiempirical models to study zinc-ligation and zinc-catalyzed processes in biological systems.

5 Development of semiempirical models for peptide hydrolysis in zinc enzymes

5.1 Abstract

Zinc is crucial for its structural, catalytic and co-catalytic roles in many proteins. The molecular details underlying the function of these enzymes can in principle be investigated using high-level quantum chemistry methods. However, ab initio and density functional theory methods are not routinely used to study reaction mechanisms of metalloproteins due to their high computational cost. Semiempirical quantum models provide a useful compromise in accuracy and computational cost. In this work, re-parameterization of the AM1 model for carbon and nitrogen is carried out for describing the nucleophilic attacks of zinc-bound hydroxyl on substrates in zinc enzymes. Previously optimized AM1 parameters for the elements H, O, and Zn are used (see Chapter 3 and Chapter 4) and parameters for the elements C and N are optimized by fitting the structural and energetic properties of a number of biomimetic complexes and transition structures. Semiempirical parameters are optimized using a genetic algorithm. The performance and average errors in comparison to B3LYP are discussed for standard AM1, PM3, and for the optimized model. Compared to AM1 and PM3 models, the optimized model yields better agreement with the benchmark B3LYP calculations. The optimized model also improves the accuracy at reproducing proton transfer energy profiles in the model complex of the active site of carbonic anhydrase. Overall results suggest that the optimized model can serve as an alternative semiempirical model to study substrate-specific and metal ligand-specific reactivity of zinc enzymes.

5.2 Introduction

Zinc is, after iron, the second most biologically abundant trace metal. Zinc deficiencies are associated with numerous health problems.[57] It is an essential co-factor of many proteins, in which it has structural, catalytic, or co-catalytic activities.[43,59,242,243] Zinc adopts a variety of coordination structures that enable a variety of biological functions.[243] At structural sites, zinc normally binds four cysteine residues or three cysteine and one histidine residues, whereas at catalytic sites it binds three histidines and one water, or two histidines, one glutamic/aspartic acid

and one water.[59,60,192] Zinc peptidases catalyze the cleavage of peptide bond via nucleophilic attacks of zinc-bound hydroxyl.[244] Zinc is characterized by a flexible coordination geometry, fast ligand exchange, Lewis acidity, and a lack of redox activity with closed d-shell. This results in Zn being highly preferred over other transition metals for catalysis in biological processes.[43]

Due to strong local electrostatic interaction and induction effect, computational investigations of zinc containing active sites using classical force field methods are very challenging. [198] Because of very high computational cost, quantum mechanical methods are not practical to apply to entire zinc enzymes. Current state-of-the-art consists in using quantum mechanical / molecular mechanical (QM/MM) where the protein active site atoms are modeled with quantum mechanical theory (often by the density functional theory (DFT)) and the rest of the protein and solvents by molecular mechanics methods (force field).[61,198,245–247] Even though this technique is generally accurate, again, due to the high computational cost of quantum mechanical methods, the QM region is often made as small as possible, which often results in losing the accurate electronic structure description of long-range interactions.[209] Semiempirical (SE) models offer a very promising alternative. SE methods introduce empirical parameters for many of the computationally expensive two-center overlap integrals and make calculations faster than conventional DFT methods. For a particular SE model, empirical parameters are obtained by fitting the various properties of the complexes derived from experiments and/or high-level calculations.

Standard SE models (e.g. AM1[73] and PM3[86]) reproduce various properties of small organic clusters[73,86,125] but they tend to give incorrect descriptions of zinc-containing compounds.[131,215] Compared to experimental results, SE models (both AM1 and PM3) show mixed performance on zinc-bound biomolecules.[248,249] Compared to high-level density functional theory, both AM1 and PM3 fail to reproduce the coordination number of sterically crowded zinc complexes.[131] AM1 and PM3 have been shown to poorly describe the energetics of carbonic anhydrase reaction mimics.[131] PM3 yields stable structures, but it overestimates the activation energy for the nucleophilic attack on CO_2.[131]

SE models are used in the study of biological systems.[118] AM1 implementation was used in QM/MM setup to study reactions in large systems including peptide hydrolysis by thermolysin.[93,250,251] The accuracy of SE models can be improved from re-parameterization while

keeping the original theoretical formulations unchanged.[133,147] In this work, SE model with AM1 Hamiltonian is re-parameterized for carbon and nitrogen to describe reactivity of zinc metalloproteins. Previously optimized AM1 parameters for hydrogen and oxygen[220] and for zinc[252] are used. The performance of the optimized model is compared with AM1 and PM3 models along with the reference results obtained from hybrid density functional theory.

5.3 Computational Methods

5.3.1 Preparation of training set and gas phase calculations

The training set is chosen so that the model reproduces structures and energetics of zinc peptidases of the form $Zn/His_3/H_2O$ (e.g. MMPs) or $Zn/His_2/Glu/H_2O$ (e.g. thermolysin) and carbonic anhydrases of the form $Zn/His_3/H_2O$. The training set includes (1) single molecules representing fragments of enzyme active sites, (2) tetrahedral zinc complexes having zinc-nitrogen and zinc-oxygen interactions, (3) hydrogen-bonded complexes between a water molecule and a proton acceptor, and (4) reactants and products complexes representing nucleophilic attacks of a zinc-bound hydroxyl on carbon dioxide and peptide substrate.

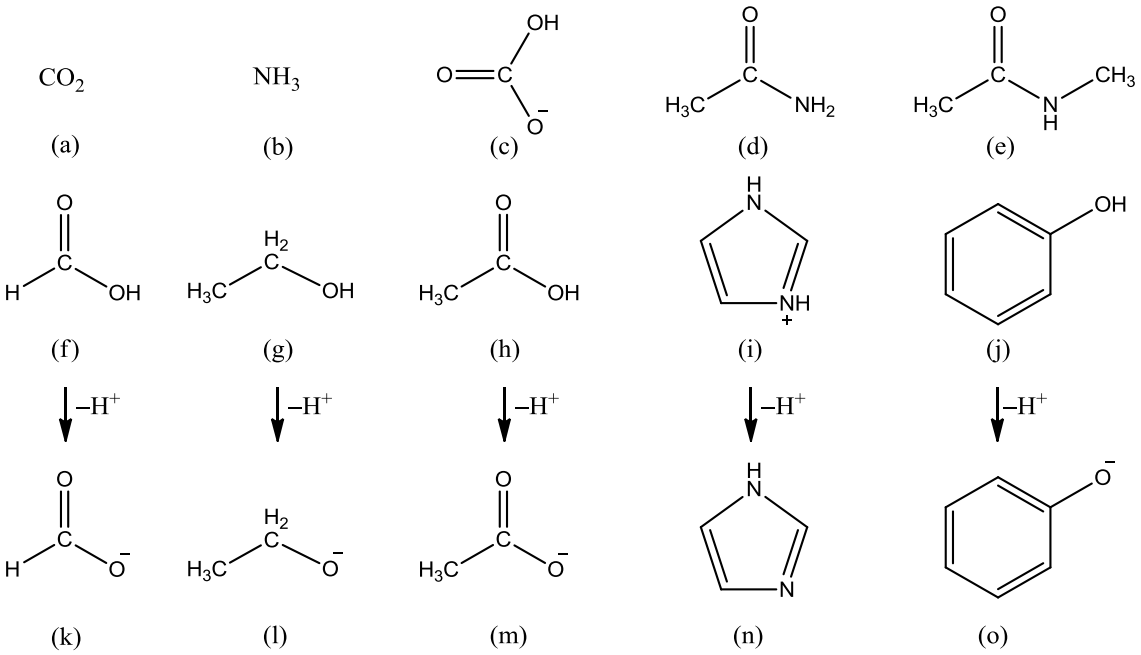


Figure 5.1: Simple compounds representing fragments of enzyme active sites. Arrows show the deprotonation of the compounds. Geometries of all compounds and proton affinities of compounds f to j are target properties to fit.

Figure 5.1 represents the model compounds of fragments mimicking the enzyme substrates, the catalytic residues involved in proton transfer, and the zinc ligands. The training set includes models for spectator ligands His (ammonia and imidazole) and Asp/Glu (formate and acetate) and models for catalytic residues involved in proton transfer with water Asp/Glu (formate and acetate), Ser (ethanolate), His (imidazole), and Tyr (phenolate). This would also include water but we do not re-parameterize water at this point, and substrates/products, which are directly involved in the OH^- nucleophilic attack (carbon dioxide, acetamide, and N-methylacetamide). For carbon dioxide to bicarbonate reaction, the whole reaction (reactants to transition states to products) is modelled whereas for acetamide and N-methylacetamide, reactants to transition states to tetrahedral intermediate are modelled.

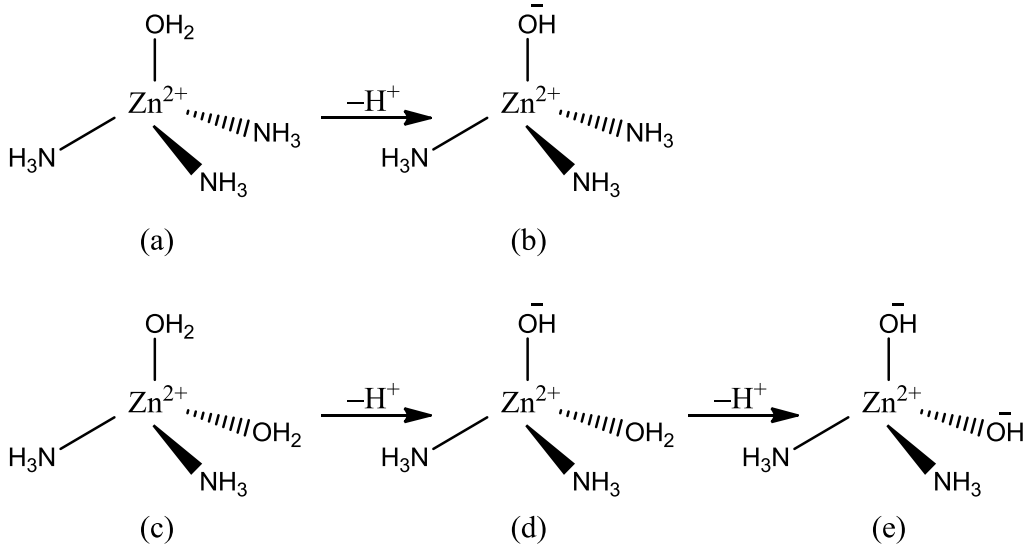


Figure 5.2: Tetra-coordinated zinc complexes mimicking metal-centered active sites of zinc enzymes, including singly (b and d) and doubly (e) deprotonated complexes. Overall geometries, Zn–N and Zn–O bond lengths, ligand-exchange reaction energies between (a) and (c), and proton affinities (shown by arrows) are target properties to fit.

To model zinc's tetrahedral coordination at active sites,[197] H_2O- and NH_3- coordinated zinc cations are included in the training set. Here, ammonia and water are used as minimal model compounds for histidine and glutamic acid. Singly and doubly deprotonated complexes are also included to model deprotonated catalytic water from zinc-bound hydroxyl (OH^-). The deprotonated water is also minimal model complex for glutamate. Figure 5.2 shows the tetrahedral compounds used in the training set with neutral, singly deprotonated, and doubly deprotonated ligands.

The first step in the catalytic cycle of zinc enzymes is the deprotonation of the zinc-bound water molecule that happens via proton transfer to a second-shell proton acceptor. To describe this step, proton transfer energy profiles from water to acetate, ethanolate, imidazole, and phenolate are also included in the training set (see Figure 5.3).

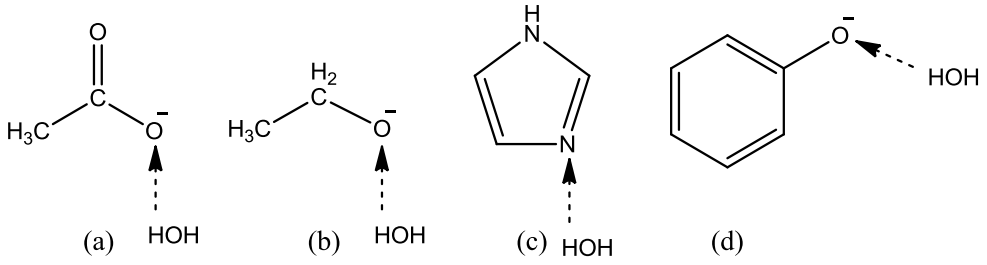


Figure 5.3: Hydrogen-bonded complexes of water and proton-accepting (a) acetate, (b) ethanolate, (c) imidazole, and (d) phenolate. Arrows represent the proton transfer. Geometries, hydrogen bonding energies, and proton transfer energies are target properties to fit.

Reactant and product complexes and reaction energies during nucleophilic attacks in zinc enzymes are also included in the training set. After deprotonation of the zinc-bound water, the zinc-bound hydroxyl performs nucleophilic attack on the carbon of peptide bond (in the case of thermolysin and matrix metalloprotease) or on the carbon dioxide substrate (in carbonic anhydrase). Figure 5.4 presents reactants and product complexes included in the training set. Carbon dioxide, acetamide, and N-methylacetamide represent the substrate for nucleophilic attacks and formate is used as a model for glutamate and aspartate.

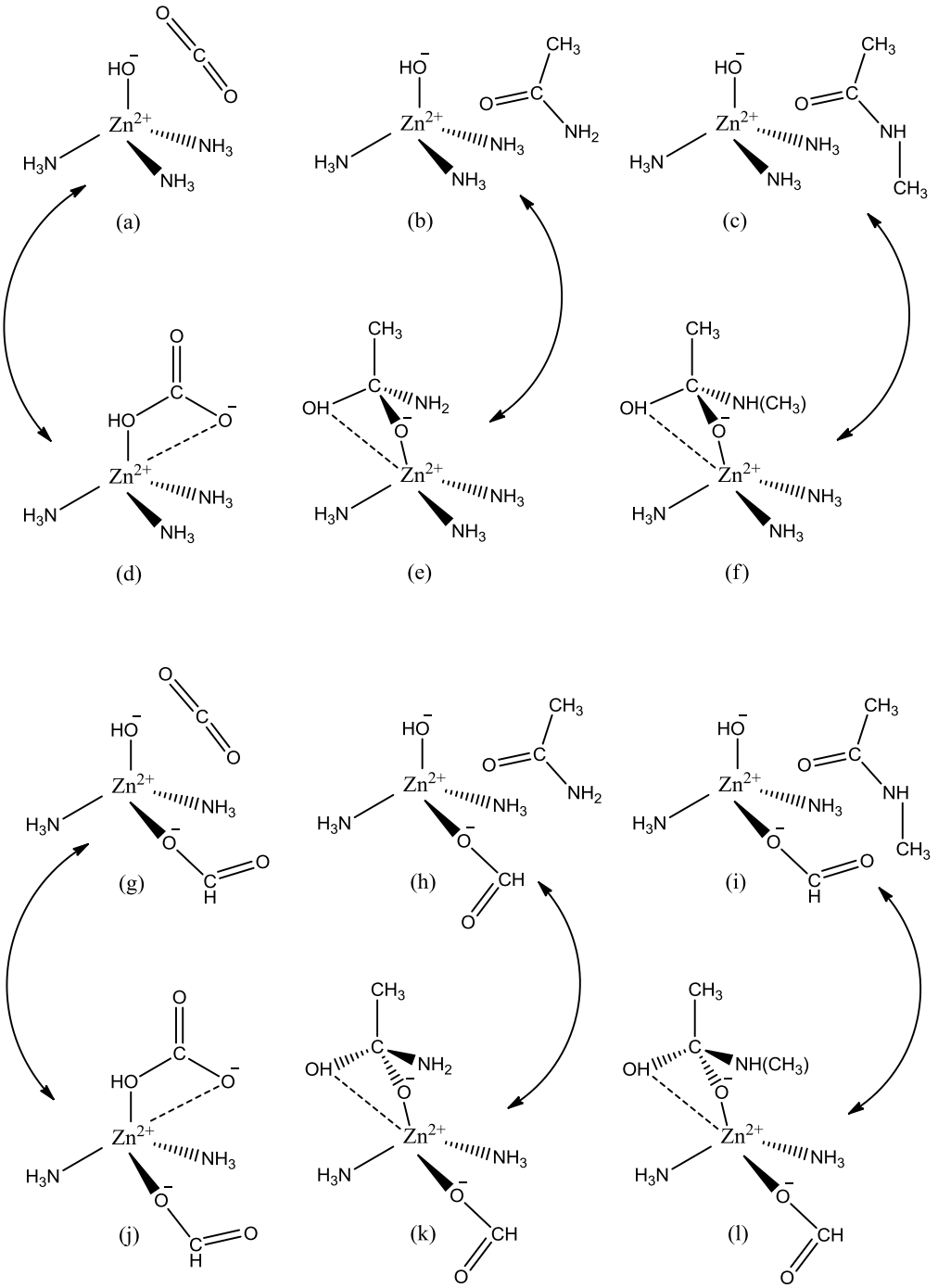


Figure 5.4: Reactant, transition-state (not shown), and product complexes representing the nucleophilic attack of a zinc-bound hydroxyl on a CO_2 substrate (a → d, g ⟶ j) and on two peptide substrates (b → e, c → f, h → k, i → l). Geometries (reactants and products), reaction energies between reactants and products, and activation energies (one from reactant to transition state and other from product to transition state) are target properties to fit.

Hybrid DFT at the B3LYP/6-311++G(d,p) level of theory was used to prepare the reference data. B3LYP is chosen as it was previously used to study reactivity in metallopeptidases.[253] Reference data were obtained from full geometry optimization Gaussian 09[150]. All transition state structures were checked using frequency calculations, to confirm that only one imaginary frequency was present. For semiempirical geometry optimization, MOPAC 07[151] was used with the optimization termination criterion GNORM = 1.0 kcal/mol/Å during the parameterization and GNORM = 0.01 kcal/mol/Å for the final assessment of the models. Cartesian coordinates of fully optimized structures from DFT calculations were used as initial coordinates for SE calculations.

To have a model describing proton transfer reactions occurring in zinc-enzymes, proton affinities of different compounds were also included in the training set. The DFT proton affinities ($E_{\mathrm{PA}}^{\mathrm{DFT}}$) and SE proton affinities ($E_{\mathrm{PA}}^{\mathrm{SE}}$) were calculated using the formulas

$$E_{\mathrm{PA}}^{\mathrm{DFT}} = -\Delta E - ZPE + \frac{5}{2}RT$$
$$E_{\mathrm{PA}}^{\mathrm{SE}} = -\Delta H + H_{\mathrm{f}}(\mathrm{H}^+) \qquad (5\text{-}1)$$

where E is the electronic energy, ZPE is the zero-point correction energy, R is the gas constant, and T is the temperature. ΔE and ΔH are the differences in electronic energy and heat of formation between the protonated and deprotonated forms of the complex. In case of SE models, the experimental heat of formation of proton [$H_{\mathrm{f}}(\mathrm{H}^+)$= 367.2 kcal/mol[221]] was used to calculate proton affinities. To model proton transfer mechanisms from zinc-bound water to second-shell proton acceptor, proton transfer profiles were also included in the training set. Energies for proton transfers were calculated by scanning the distance from H (on the proton-donating water molecule) to O/N (on the proton-accepting residue) from 1.7 to 0.9 Å in steps of -0.2 Å while keeping O (donor water molecule), H, and O/N (acceptor residue) atoms collinear (see Figure 5.3). Five points along the proton transfer energy profiles were calculated. For each point, the proton transfer energy, $E_{\mathrm{PT}}(k)$ is calculated as

$$E_{\mathrm{PT}}(k) = E(k) - E(0) \qquad (5\text{-}2)$$

where k (= 1 to 5) is the each point along the proton transfer scan, $E(k)$ is the energy of k^{th} complex, and $E(0)$ is the energy of reactant complex (hydrogen-bonded water and proton acceptor complex with no constraint).

Hydrogen bonds play crucial roles in many enzymes.[254] They often contribute to the overall stability of the active site and to the stabilization of the transition state. They also stabilize the configurations of proton-donating and proton-accepting moieties most favorable to proton transfer reactions. Special attention was given to the hydrogen bonding between water and proton acceptors by adding hydrogen-bonding energies to the training set (see Figure 5.3). Hydrogen bonding energy between molecule A and water (W) is calculated as

$$E_{\text{H-bonding}} = E_{\text{A}} + E_{\text{W}} - E_{\text{A-W}} \tag{5-3}$$

where E_{A}, E_{W}, and $E_{\text{A-W}}$ are energies of isolated molecule A, isolated molecule W, and hydrogen-bonded complex A–W.

5.3.2 Error function and parameterization procedure

To train the SE model, the error function χ is defined as the sum of errors (deviations from the DFT results) on each of the properties included in the training set. The original AM1 parameters for C and N were re-optimized by minimizing the error function. It is defined as

$$\begin{aligned}\chi \;&= W_{\text{S}} \sum_i |MSD_i| + W_{\text{B}} \sum_j \left| B^{\text{model}}(j) - B^{\text{ref}}(j) \right|^2 \\ &+ W_{\text{R}} \sum_i \left| E_{\text{R}}^{\text{model}}(i) - E_{\text{R}}^{\text{ref}}(i) \right| + W_{\text{PA}} \sum_i \left| E_{\text{PA}}^{\text{model}}(i) - E_{\text{PA}}^{\text{ref}}(i) \right| \\ &+ W_{\text{PT}} \sum_i \left[\frac{1}{n_{\text{PT}}} \sum_{k=1}^{n_{\text{PT}}} \left| E_{\text{PT}}^{\text{model}}(i,k) - E_{\text{PT}}^{\text{ref}}(i,k) \right| \right] \\ &+ W_{\text{TS}} \sum_i \left| E_{\text{TS}}^{\text{model}}(i) - E_{\text{TS}}^{\text{ref}}(i) \right| + W_{\text{HB}} \sum_i \left| E_{\text{HB}}^{\text{model}}(i) - E_{\text{HB}}^{\text{ref}}(i) \right| \\ &+ W_{\text{LE}} \sum_i \left| E_{\text{LE}}^{\text{model}}(i) - E_{\text{LE}}^{\text{ref}}(i) \right|\end{aligned} \tag{5-4}$$

where W_S (= 3×10^3 kcal/mol/Å^2) is the weighting factor for mean square deviations of structures (MSD), W_B (= 1.5×10^3 kcal/mol/Å^2) is the weighting factor for bond lengths (B), W_R (= 0.6×10^2 kcal/mol) is the weighting factor for reaction energies (E_R) and W_{PA} (= 10^2 kcal/mol) is the weighting factor for proton affinities (E_{PA}), W_{PT} (= 5×10^2 kcal/mol) is the weighting factor for proton transfer energies (E_{PT}), W_{TS} (= 0.75×10^3 kcal/mol) is the weighting factor for activation energies (E_{TS}), W_{HB} (= 7×10^2 kcal/mol) is the weighting factor for hydrogen-bonding energies (E_{HB}), and W_{LE} (= 7×10^2 kcal/mol) is the weighting factor for ligand exchange energies (E_{LE}). The units of weighting factors were chosen so that the error function is in kcal/mol. $n_{PT} = 5$ is the number of points along each proton transfer profile. Superscript "model" represents SE and superscript "ref" represents DFT. Index j(=1 to 20) represents the Zn···N/O bond distances in the tetra-coordinated zinc complexes (four bond distances for each of the complexes in Figure 5.2). i is the index of the compound or cluster for which the property is calculated. The error function contains 118 terms in total: 46 mean square deviations in structures, 20 Zn-N/O bond lengths, 6 reaction energies (difference in energy between reactant and product), 8 proton affinities, 12 activation (transition state) energies, 20 proton transfer energies, 4 hydrogen bonding energies, and 2 ligand exchange energies (NH_3 to H_2O exchange from Figure 5.2 (a) to Figure 5.2 (c) and HCO_3^- to H_2O exchange from Figure 5.4 (d) to Figure 5.2 (a)). Properties included in the training set and corresponding number of terms by the figure numbers are presented in Table 5.1.

Table 5.1: Properties and numbers of terms included in the error function. For clarity, numbers are broken down by Figures.

Properties	Number of terms	Numbers of terms in each Figure			
		Fig. 5.1	Fig. 5.2	Fig. 5.3	Fig. 5.4
MSD	46	15	5	8	18
Bond length	20		20		
Reaction energy	6				6
Proton affinity	8	5	3		
Activation energy	12				12
Proton transfer energy	20			20	
Hydrogen bonding energy	4			4	
Ligand exchange energy	2		1*		1**
Total	118	20	29	32	37

* NH_3 to H_2O exchange corresponding to a transition from complex Figure 5.2 (a) to Figure 5.2 (c)

** HCO_3^- to H_2O exchange corresponding to a transition from Figure 5.4 (d) to Figure 5.2 (a)

The parameterization procedure is followed as described previously.[120] PIKAIA, a genetic algorithm, was used to optimize the SE parameters. For each round of optimization, a population of 100 individuals was evolved for 300 generations. Each individual, within the search region, represents a set of SE parameters. For each set of parameters, fitness is defined as $1/(\chi + 1)$. The set of parameters that has the highest fitness score is considered as the final set of parameters in each run. Previously optimized AM1 parameters for hydrogen,[220] oxygen,[220] and zinc[252] were used for elements H, O, and Zn during re-parameterization of AM1 parameters for C and N. Since nitrogen atoms directly coordinate zinc, AM1 parameters for nitrogen were optimized first. Carbon parameters were optimized second, using the nitrogen parameters optimized previously. Original AM1 parameters[73] for both carbon and nitrogen were used as initial parameters and

allowed a ±50% change from their original values during the optimization. Three rounds of optimization (nitrogen, then carbon) were needed to converge the error value. To minimize the error value further, the optimization was repeated by gradually reducing the search region, allowing only ±20%, ±10%, ±5%, ±4%, ±3%, ±2%, and ±1% change from the previous optimal parameters set. For ±5% to ±1% search regions, only two rounds of parameterization were needed to minimize the error value. Parameterization was terminated after no improvement was seen over one full round of optimization in the ±1% search region. In this work, 24 AM1 parameters for carbon and 21 AM1 parameters for nitrogen were optimized.

5.4 Results and discussion

The optimized AM1 parameters for C and N are presented in Table 5.2. Previously optimized parameters for H, O, and Zn are also presented.[220,252] Results from the optimized model are compared with AM1 and PM3 (along with B3LYP values) for both the training and the testing sets. To evaluate the performance of the optimized model on zinc enzymes, energy profiles for proton transfer of the complexes that mimic the active sites of zinc-enzymes (e.g. thermolysin and carbonic anhydrase) are also calculated and results are compared with AM1 and PM3 (along with B3LYP values).

Table 5.2: Optimized AM1 parameters for C and N along with previously optimized H, O, and Zn.[220,252] Symbols have their usual meaning.[73]

Parameters	**H**	**C**	**N**	**O**	**Zn**
U_{ss} (eV)	–12.560002	–32.150200	–60.031992	–124.312581	–18.409395
U_{pp} (eV)		–28.967801	–56.610516	–79.588146	–14.386244
ζ_s (au)	1.078537	1.746775	1.652061	3.497468	2.170903
ζ_p (au)		1.498788	2.218407	2.528886	1.320255
β_s (eV)	–4.860006	–10.277864	–14.706789	–37.859064	–1.249800
β_p (eV)		–14.119788	–17.473830	–37.180813	–5.141664
G_{ss} (eV)	14.331944	8.024315	11.578123	21.583836	8.305369
G_{sp} (eV)		7.448147	10.555559	18.193541	8.193452
G_{pp} (eV)		5.276066	14.729169	17.042124	13.641224
G_{p2} (eV)		5.942355	11.849208	10.338310	12.571464
H_{sp} (eV)		2.649657	4.521688	2.811269	0.589834
α (Å^{-1})	2.739361	2.522522	2.987975	5.799824	1.498274
K_1 (eV)	0.049143	0.018731	0.019546	0.352574	
L_1 (Å^{-1})	3.953000	5.451509	4.891240	3.030000	
M_1 (Å)	0.980400	1.750632	1.864377	0.974529	
K_2 (eV)	0.003321	0.039920	0.031066	0.048729	
L_2 (Å^{-1})	8.256000	6.889123	4.023180	6.150900	
M_2 (Å)	1.345680	2.112193	2.086714	1.841512	
K_3 (eV)	–0.021350	–0.026641	–0.004588		
L_3 (Å^{-1})	1.897200	4.662174	2.706654		
M_3 (Å)	1.270080	1.782816	2.879079		
K_4 (eV)		–0.000826			
L_4 (Å^{-1})		3.168583			
M_4 (Å)		2.215877			

5.4.1 Performance on the training set

Figure 5.5 compares the performance of AM1, PM3, and the optimized model (this work) on the training set. As shown in Figure 5.5 (a) and Table 5.3, the average mean square deviation in structures from the optimized model (0.16 $Å^2$) is smaller than AM1 (0.40 $Å^2$) and PM3 (0.45 $Å^2$). In addition, the average unsigned error (AUE) for Zn–N and Zn–O bond lengths suggests that the optimal model depicts the B3LYP geometries better than AM1 and PM3.

As Table 5.3 shows, the optimized model improves the accuracy in energies for proton transfer, hydrogen bonding, and ligand exchange more than for reaction energies, proton affinities, and activation energies. The optimized model gives an AUE of 5.14 kcal/mol for proton transfer energies, compared to 18.58 kcal/mol for AM1 and 12.12 kcal/mol for PM3. Moreover, the optimized model captures the hydrogen bonding energies of the systems shown in Figure 5.3. The optimized model yields an AUE of 1.40 kcal/mol for hydrogen bonding energies, compared to 3.02 kcal/mol for AM1 and 2.17 kcal/mol for PM3. For ligand exchange reactions, both AM1 (14.78 kcal/mol) and PM3 (17.32 kcal/mol) give higher absolute errors in ligand exchange energy than the optimized model (0.50 kcal/mol).

During proton transfer scanning in water-phenolate and water-imidazole pairs, MOPAC fails to correctly optimize the structures with H(donor)···O (acceptor) for phenolate and H(donor)···N(acceptor) for imidazole constraints of 1.5, 1.3, 1.1, and 0.9 Å. This results in the wrong structures, and hence results are not reported for all the models. For these two pairs, only 1.7 Å constraint results are shown in Figure 5.5 (d). The optimized model also produces large errors in proton transfer for water-acetate pair for H (donor)···O (acceptor) constraint of 0.9 Å (see Figure 5.5 (d)) due to the fact that methyl group from acetate donates proton to OH^- (of proton donating water) to form H_2O. AM1 and PM3 fail to optimize the $[Zn^{2+}\cdot OH^-\cdot 2NH_3\cdot HCOO^-$ ··· N-methylacetamide] (see Figure 5.4 (i)) complex during the optimization for transition state, and hence activation results are not reported.

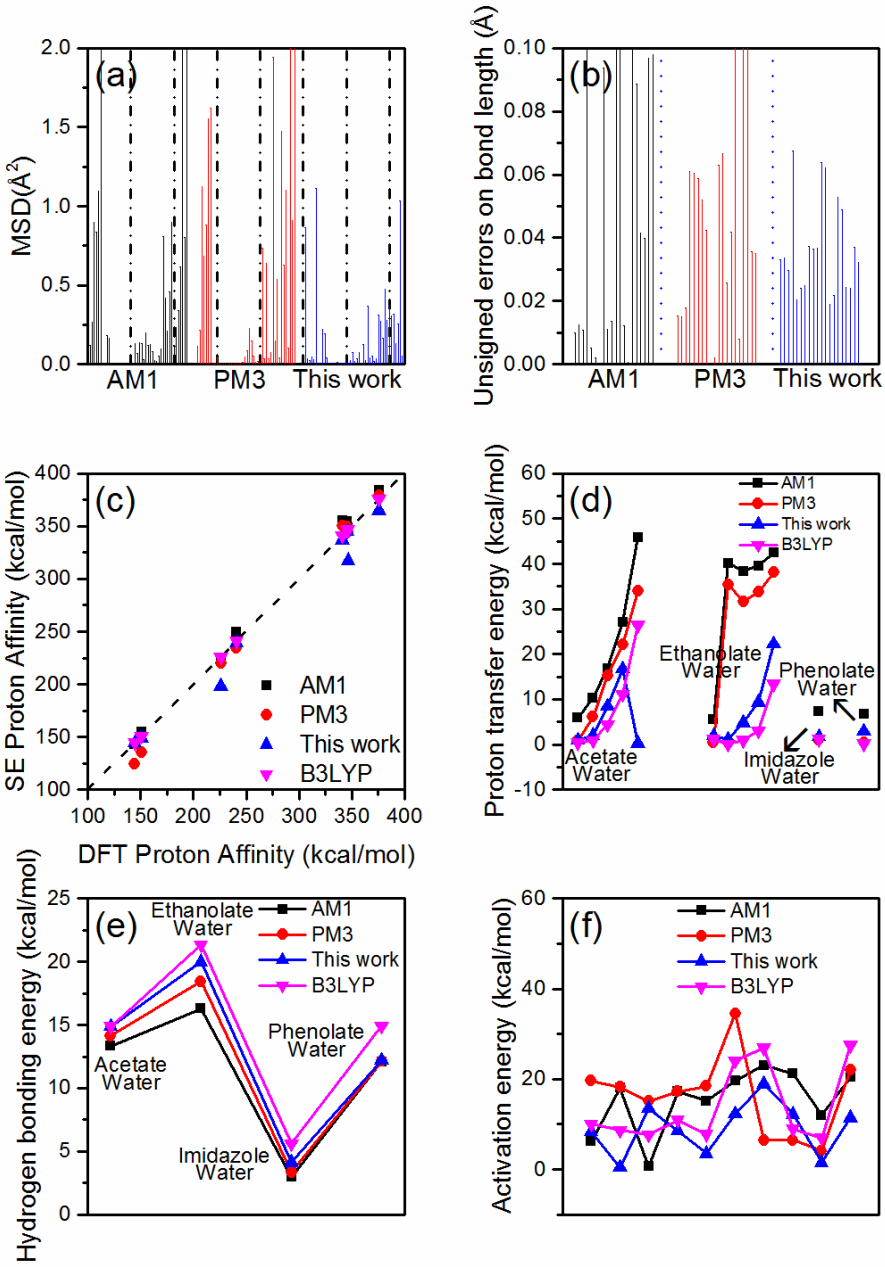


Figure 5.5: Performance of AM1 (black), PM3 (red), and the optimized model (blue) on the training set in comparison with B3LYP (pink) results. Average unsigned errors are shown for (a) mean square deviation (*MSD*) in structures and (b) Zn–N and Zn–O bonds lengths from tetrahedral complexes shown in Figure 5.6. The horizontal axes for figures (d) and (e) are hydrogen-bonded water-ligand complexes for proton transfer profile energies and hydrogen bonding energies. The horizontal axis for figure (f) is the complexes index in activation energy calculations as shown in Figure 5.4.

Table 5.3: Average unsigned errors of different properties in the training set.

Properties	Average unsigned errors		
	AM1	PM3	This work
Training set			
MSD ($Å^2$)	0.40	0.45	0.16
Bond length (Å)	0.05	0.05	0.03
Reaction energy (kcal/mol)	12.89	5.87	9.30
Proton affinity (kcal/mol)	6.50	8.46	9.49
Activation energy (kcal/mol)	6.67	8.82	6.69
Proton transfer energy (kcal/mol)	18.58	12.12	5.14
Hydrogen bonding energy (kcal/mol)	3.02	2.17	1.40
Ligand exchange energy (kcal/mol)	14.78	17.32	0.50
Testing set			
Bond length (Å)	0.08	0.04	0.02
Proton affinity (kcal/mol)	4.64	7.40	2.53
Activation energy (kcal/mol)	2.82	5.20	3.85

5.4.2 Performance on the testing set

AM1 parameters for C and N were obtained by fitting various properties of small clusters representing the active sites of zinc enzymes. The transferability of the optimized model is assessed for large systems, which are more representative of the active sites of zinc enzymes. As shown in Figure 5.6, two tetra-coordinated structures (three zinc-bound imidazoles and one water; and two imidazoles, one acetate, and one water) representing the active sites of carbonic anhydrase and thermolysin are considered. The nucleophilic attacks on carbon dioxide, acetamide, and N-methylacetamide are studied by considering the reactants, transition states, and tetrahedral intermediates as shown in Figure 5.8.

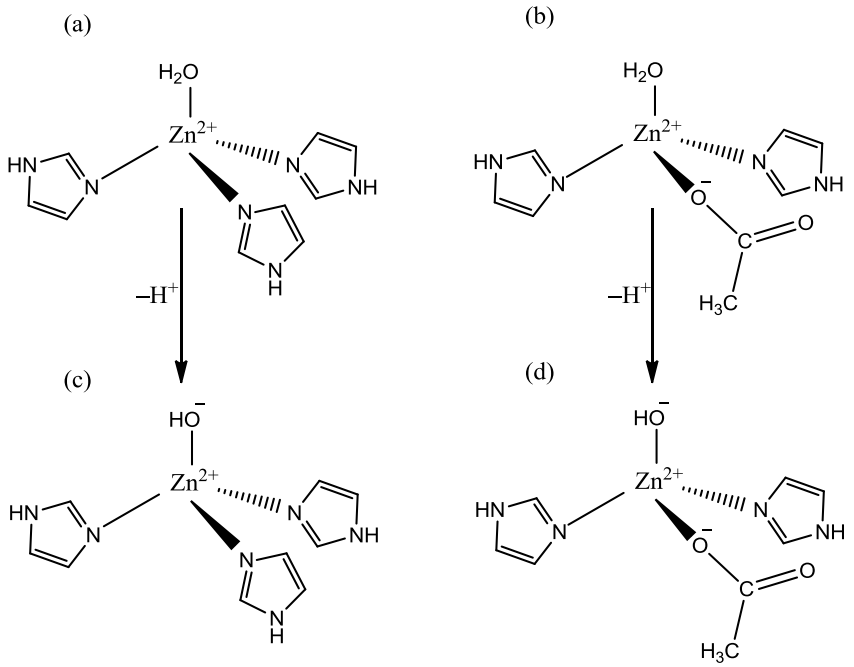


Figure 5.6: Tetra-coordinated zinc complexes representing metal-center complexes of carbonic anhydrase (a) and thermolysin (b) and their deprotonated forms in (c) and (d).

As shown in Figure 5.7, AM1 gives large errors in Zn–O and Zn–N bond lengths with an AUE of 0.08 Å (see Table 5.3). Original AM1 parameters for Zn were optimized by fitting heats of formation, ionization energies, dipole moments, and geometries of organozinc compounds such as zinc with alkyl, halides, and alkyl-halide mixed compounds.[222] Our complexes and properties do not fall within the scope of the original parameterization of zinc, hence original AM1 model fails to reproduce Zn–O/N bond lengths of the complexes used in this work. Moreover, AM1 stabilizes the acetate ligand of Figure 5.6 (b and d) into a bidentate form inconsistent with the B3LYP optimized geometry. PM3, however, improves the accuracy on Zn–O and Zn–N bond lengths with an average unsigned error of 0.04 Å (see Table 5.3). Improving on the performance of AM1 and PM3, the optimized model yields an AUE of 0.02 Å for Zn–O and Zn–N bond lengths. The optimized model also offers improved performance over AM1 and PM3 at reproducing proton affinities of larger complexes.

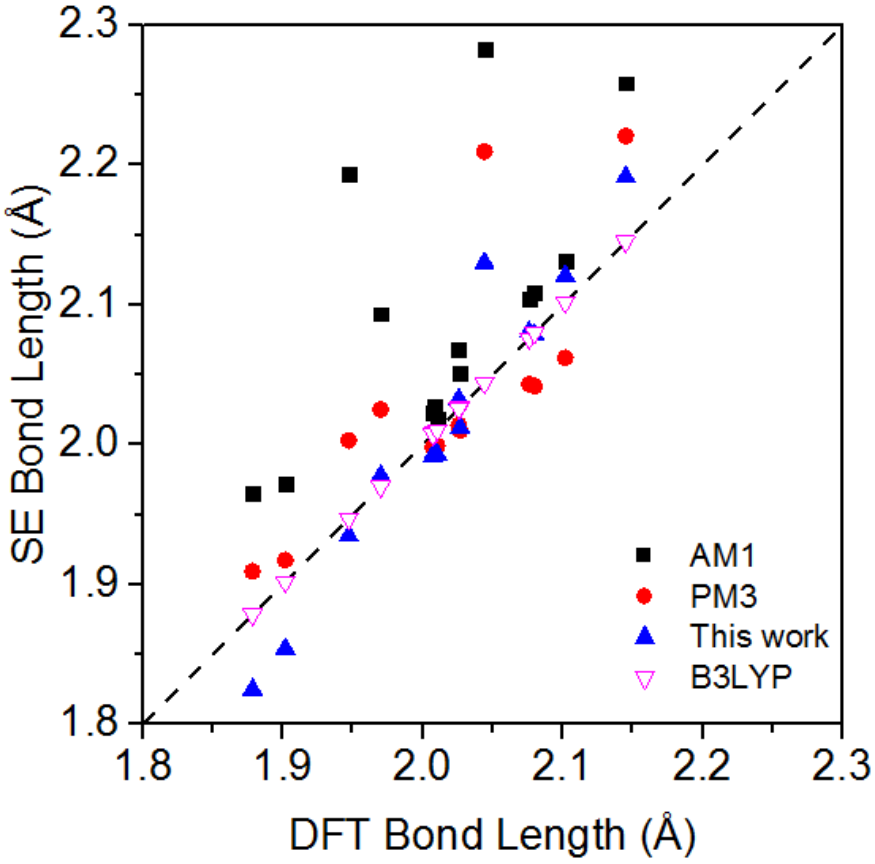


Figure 5.7: Zn–O and Zn–N bond lengths in the tetra-coordinated zinc complexes shown in Figure 5.6. The diagonal dashed line shows perfect correlation.

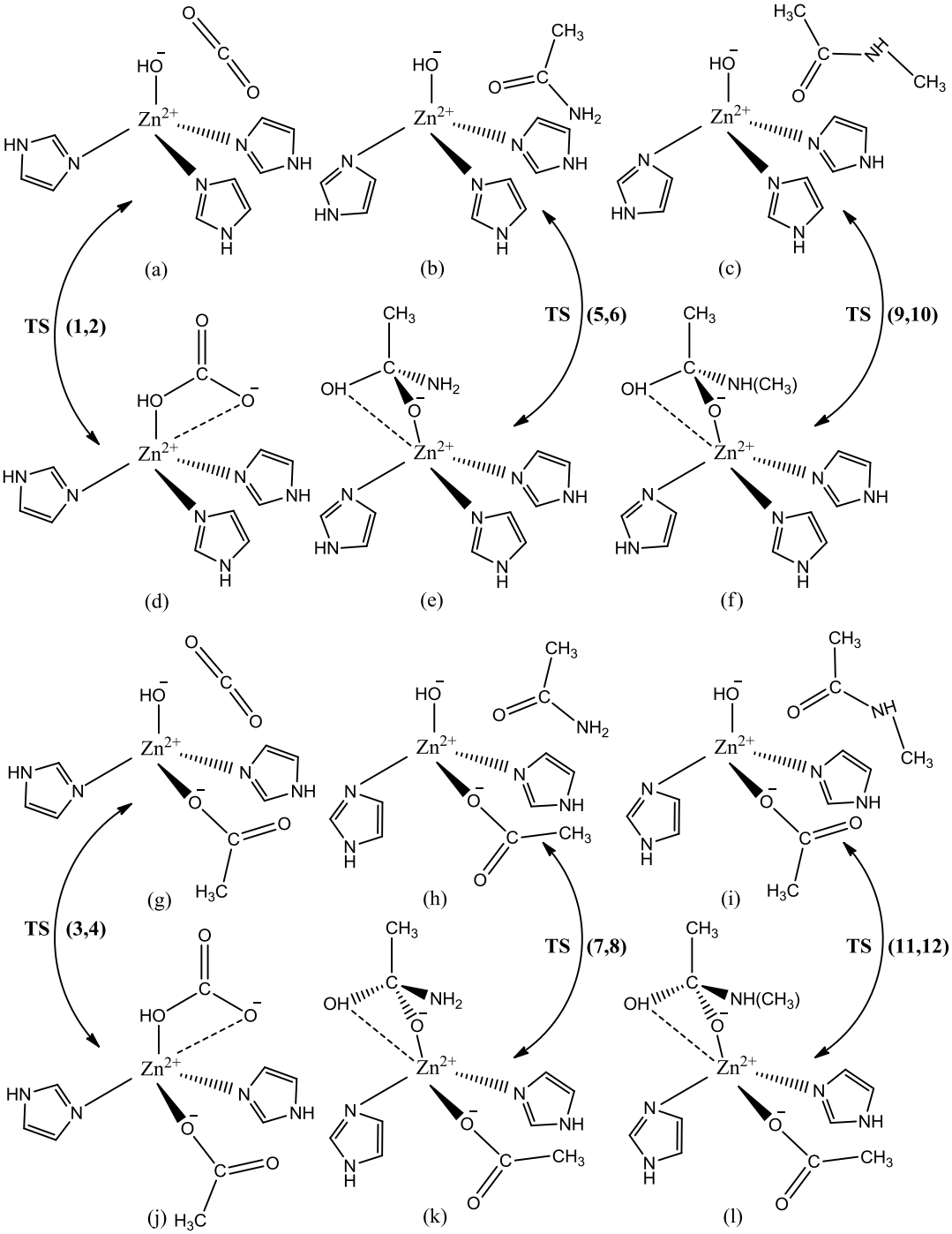


Figure 5.8: Testing set. Testing set contains the model complexes (reactants, transition states (TS), and product/tetrahedral intermediate) for carbonic anhydrase ((a) and (d)), matrix metalloproteinases ((b)-(e) and (c)-(f)), and thermolysin ((h)-(k) and (i)-(l)). Numbers 1 to 12 represent the activation energies associated with each complex.

Activation energies, from both the reactant state and the product state, were calculated for the structures shown in Figure 5.8. Similarly to the training set (see Figure 5.4)), the testing set includes nucleophilic attacks on carbon dioxide, acetamide, and N-methylacetamide. The absolute errors in activation energy are presented in Figure 5.9. Figure 5.9 and Table 5.3, AM1 gives a low AUE in the activation energy of 2.82 kcal/mol whereas PM3 gives an AUE of 5.20 kcal/mol. The optimized model, however, gives an AUE of 3.85 kcal/mol for activation energies, which lies between the AM1 and PM3.

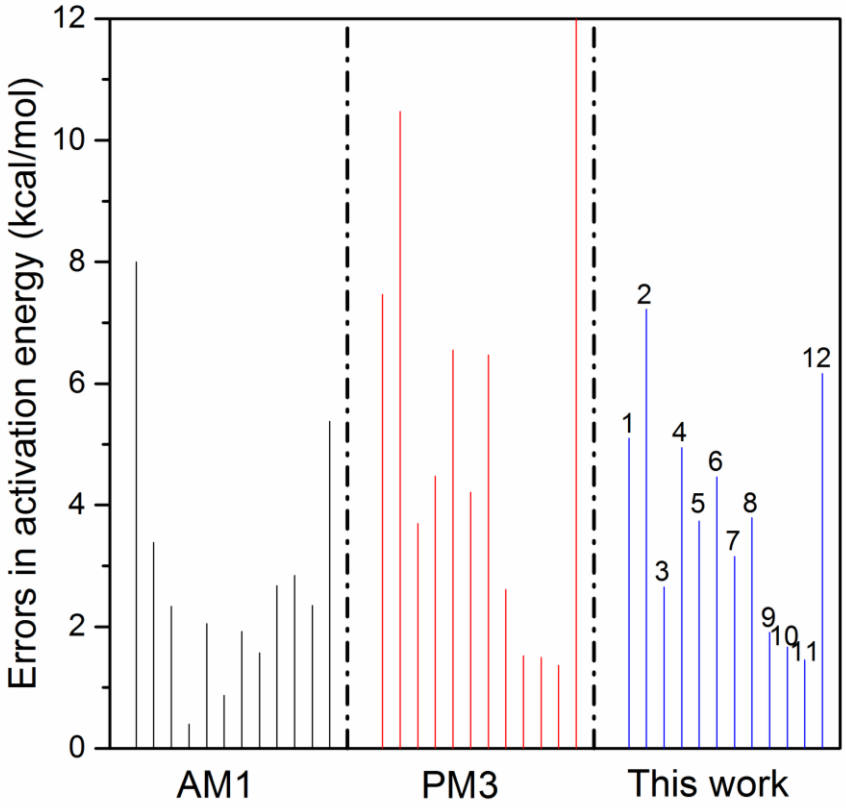


Figure 5.9: Comparison of the absolute errors in activation energies from AM1, PM3, and the optimized model in reference with B3LYP. Numbers 1 to 12 represent the errors in activation energies coming from the respective complexes shown in Figure 5.8. Same numbering applies for AM1 and PM3.

5.4.3 Proton transfer energy profiles in thermolysin and carbonic anhydrase gas-phase models

Thermolysin and carbonic anhydrase are relatively small and widely studied zinc enzymes. These enzymes can be used as model complexes for many biological systems. [61,245,255,256]

To test the optimized model on larger structures more representative of the complete active site of a zinc enzyme, proton transfer energy profiles were calculated in structures mimicking thermolysin and carbonic anhydrase active sites (see Figure 5.10 (a) and Figure 5.11 (a)). Proton transfer energies, relative to the minimum-energy configuration, were calculated by scanning the H(donor) and O(acceptor) distances while keeping O(donor), H(donor), and O(acceptor) atoms collinear. Results are compared with AM1 and PM3 in reference with B3LYP results. Figure 5.10 (a) presents the model active sites for thermolysin and Figure 5.11 (a) for carbonic anhydrase. Active site coordinates were taken from high-resolution crystal structures of thermolysin (PDB code: 1LNF) and carbonic anhydrase (PDB code: 2CBA). For both systems, hydrogen atoms were generated using CHARMM-GUI.[257] Positions of the hydrogen atoms and zinc-bound water molecule were optimized at the B3LYP/ 6-31G(d) level of theory while keeping all other heavy atoms fixed at their original crystallographic positions. Energies were calculated by performing rigid scan calculations at B3LYP/6-311++G(d,p) level, with all atoms kept at their initial positions and only the H atom being moved. DFT optimized structures were used for all SE calculations. For DFT and SE methods, rigid scan calculations were performed with Gaussian 09[150].

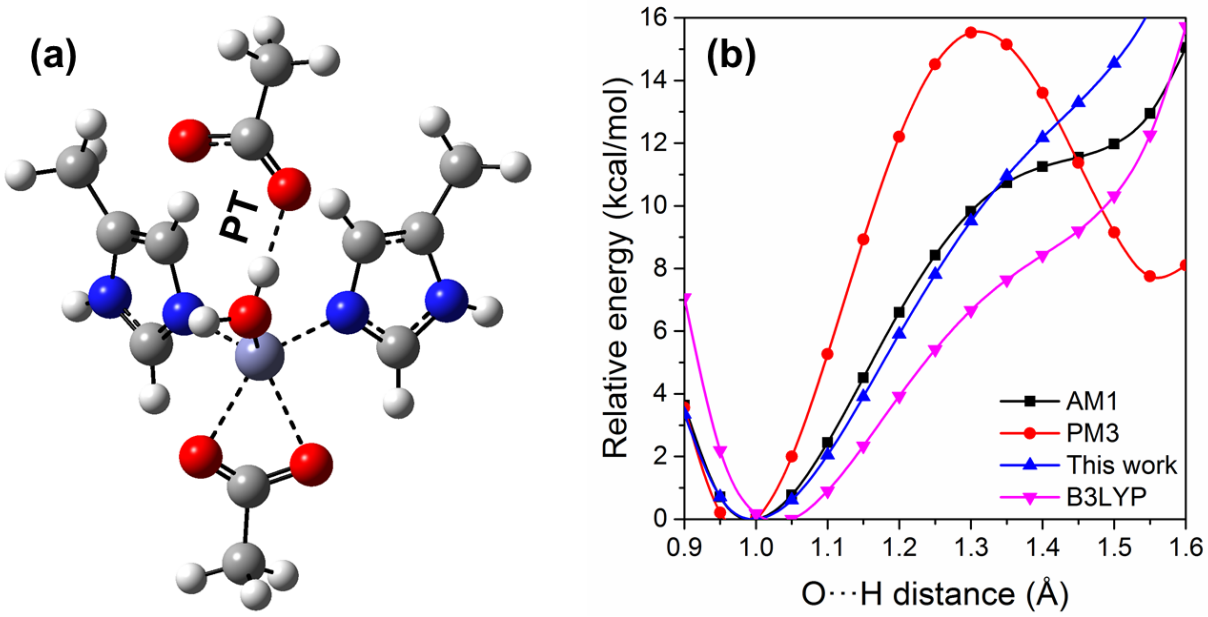


Figure 5.10: (a) Cluster model of the active site of thermolysin in which zinc is coordinated by two histidines, one glutamate, and one water molecule. PT represents the proton transfer from zinc-bound water to proton acceptor glutamate. Glutamates are modeled with acetate and histidines are modeled with 5-methylimidazole. (b) Proton transfer energies profiles calculated from AM1, PM3, the optimized model (this work), and B3LYP by constraining O(donor) and H(donor) and by keeping O(donor), H(donor), and O(acceptor) collinear.

Figure 5.10 (b) presents the SE energy profiles for proton transfer from zinc-bound water to acetate (model complex for glutamate) at the active site of thermolysin as compared to B3LYP energy profile. The PM3 model severely overestimates the proton affinity of the carboxylate moiety. The AM1 model does not display a stable protonated acetate structure like PM3 but it has a noticeable shoulder around 1.4-1.5 Å. The average unsigned error, in reference to B3LYP, is 1.98 kcal/mol for AM1, 4.87 kcal/mol for PM3, and 2.57 kcal/mol for the optimized model.

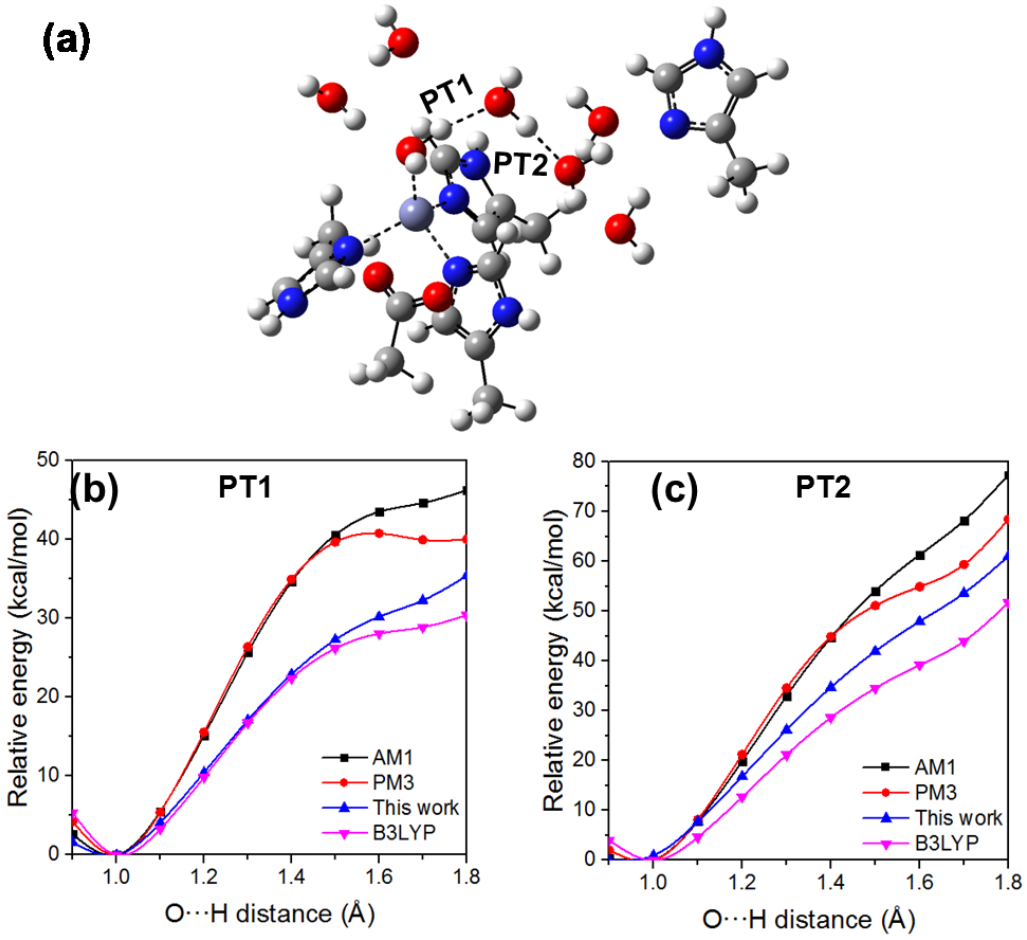


Figure 5.11: (a) Cluster model of the active site of carbonic anhydrase. Complex representation is same as in Figure 5.10. PT1 represents the proton transfer from zinc-bound water to the next second-shell water molecule and PT2 represents the proton transfer between two water molecules within the hydrogen-bonded water chain. (b) Energy for proton transfer of PT1.(c) Energy for proton transfer of PT2. Results from AM1, PM3, and the optimized models are compared with B3LYP.

In carbonic anhydrase, proton transfer occurs via a chain of hydrogen-bonded water molecules.[245] As shown in Figure 5.11(a), two proton transfer paths, PT1 and PT2 are considered. Figure 5.11(b) and (c) present, energy profiles for proton transfer for PT1 and PT2. In both cases, AM1 and PM3 give similar energy profiles. Both models overestimate the relative proton transfer energies with average unsigned errors of 9.28 kcal/mol (AM1) and 7.80 kcal/mol (PM3) for PT1 and 13.33 kcal/mol (AM1) and 10.81 kcal/mol (PM3) for PT2. Unlike AM1 and PM3, the

optimized model reproduces relative proton transfer energies in excellent agreement with B3LYP for PT1. Despite overestimating relative energies in PT2, the optimized model still shows an improvement in accuracy compared to AM1 and PM3. The average unsigned errors in proton transfer from the optimized model are 1.76 kcal/mol for PT1 and 5.84 kcal/mol for PT2. The overall results suggest that the optimized model can be used to explore proton transfer mechanisms in zinc enzymes.

5.5 Conclusion

A semiempirical model based on AM1 Hamiltonian was re-parameterized for carbon and nitrogen to describe peptide hydrolysis in zinc enzymes. Previously optimized AM1 parameters for hydrogen and oxygen (developed to describe proton transfer in water), and zinc (for zinc-bound water and hydrogen sulfide) were used during the optimization of AM1 parameters for C and N. A new set of parameters is obtained by minimizing the error on a number of properties important to predict the reactivity of the zinc enzyme: coordination geometries, proton transfer energies, and formation energies of the tetrahedral intermediates or the products.

Compared to AM1 and PM3, the optimized model improves the accuracy in overall structures of complexes including tetrahedrally coordinated systems, hydrogen bonding, and ligand exchange reactions in reference to DFT results. Moreover, the test results on gas-phase models suggest that the optimized model can be used to study proton transfer mechanisms in zinc-enzymes. The optimized model can also be used in full quantum mechanics and hybrid quantum mechanics/molecular mechanics molecular dynamic simulations to explore various reactions in zinc proteins. This parameterization technique can be used to develop new reaction-specific SE models, or more complex SE models with alternate core-repulsion functions. The performance of the SE models can also be improved by developing new models including *d*-orbitals.

6 Conclusions and outlook

6.1 Conclusions

Theoretical modelling of chemically active biomolecules is a challenging problem that requires state-of-the-art techniques from computational physics, and considerable chemical and biological insight. Semiempirical quantum methods provide a promising alternative to ab initio methods, and enable the study of such systems with acceptable accuracy and at reasonable computational cost. Due to overly simplified Hamiltonian, the accuracy of semiempirical models is poor for complexes and reactions that fall outside the scope of their development. Reaction-specific parameterization technique improves the accuracy of semiempirical models in structures and energies of specific classes of systems.

Despite having many semiempirical (SE) models and approximate density functional theory for water,[124,145,146,258,259] to the best of our knowledge, no model describes both proton transfer mechanisms and the static and dynamic properties of liquid water with an accuracy comparable to ab initio methods. The AM1-LW model we have developed improves the accuracy and yields structural and thermodynamic properties of water in line with available high-level theoretical and experimental results. AM1-LW model also yields very sensitive properties of water such as gas-phase vibrational frequencies, dielectric properties, and infrared spectrum in good agreement with experiment.

The optimized SE model for zinc and sulfur, using AM1-LW model for hydrogen and oxygen, optimized for proton transfer in water, reproduces structural and energetic gas-phase properties of zinc-bound water and zinc-bound hydrogen sulfide. The model also reproduces the hydration structure of the zinc cation and the solvation structure of zinc in liquid hydrogen sulfide. The model is further used to estimate the pK_a shift of zinc-bound water and zinc-bound hydrogen sulfide in liquid water. The estimated zinc-induced pK_a shifts suggest that zinc increases the acidity of ligand in good agreement with experimental and theoretical values.

Using previously optimized parameters for hydrogen, oxygen, and zinc in this work, the optimized SE model for carbon and nitrogen consistently reproduces the structure and energetics of elementary reactions catalyzed by zinc enzymes, such as the proton transfer from a zinc-bound

water molecule to a proton-accepting glutamate, or the nucleophilic attack of a peptide substrate by a hydroxide ion.

In this work, new reaction-specific semiempirical models were developed that show significant improvement at reproducing gas-phase and liquid-phase structural and energetic properties of hydrogen-bonded biomolecules. The models can also be very useful to explore different mechanisms in zinc metalloproteins. Especially one can explore the proton transfer paths and influences of the neighbor ligands to the active sites of the zinc enzymes. The optimized model shows great improvement over existing standard semiempirical models, and gives a promising outlook for its application to larger biomolecular systems. The current work suggests that more targeted parameterization approaches can improve the inherent limitations of semiempirical models. For systems for which a given SE model has shortcomings, separate models should be developed. Once many different semiempirical models have been developed for a variety of reactions/compounds, new models can easily be developed using the genetic algorithm approach, by mixing old models into the population. Thus, their most useful "traits" can be selected.

6.2 Outlook

The main advantage of semiempirical quantum methods is their low computational cost compared to other high-level quantum mechanical methods such as density functional theory. Because semiempirical models use many approximations, accuracy and transferability are significant concerns. To address these limitations of the SE models, different techniques, such as re-parameterizing existing SE models, introducing various correction terms, or larger basis sets, are used with or without changing the original mathematical formulations. Improving SE models by changing core-repulsion functions and re-parameterizations of previously developed models for reaction-specific systems often reduce transferability concerns. New SE models can be further developed for variety of physical, chemical, and biological systems. The techniques, used in the thesis, may also be used to design and study reactivity of solids, polymers, biomaterials, etc. These models can also be tested for their application in modeling nanomaterials (e.g. metal-organic frameworks) for biomedical applications.[260,261]

One can apply the optimized models in molecular dynamic simulations at full SE QM or SE QM/MM levels to investigate long-range interactions and long time-scale properties of biologically relevant systems. Moreover, thanks to their high computational efficiency, these models can be used to compute free energy profiles of catalytic reactions, which often require extensive conformational sampling. The following studies can be done using the optimized models from this work.

1. As hydrogen bonds play key roles in water, the AM1-LW model can be used to study roles of hydrogen bonds in the diffusivity of water. The study can further be extended to understand forming and cleavage of hydrogen bonds during reorientation of water. The model can be used to study translational and rotational dynamics[262] of water to understand origin of some of its anomalies. AM1-LW model can also be used to study solvent dynamics around the dissociation products of water (H^+ and OH^-) as they form and recombine into H_2O molecules. The effect of the ions or solute during the recombination processes can also be investigated.
2. The optimized models can be used to study the preferential coordination of zinc by various ligands in aqueous solutions. Molecular dynamic simulations can be performed on zinc cation solvated in a mixture of different ligands (e.g., water, ammonia, and hydrogen sulfide) to see which ligand combinations are preferred by zinc. The results would suggest which ligand arrangements are most selective to zinc ions, and provide insight on how proteins selectively capture zinc ions. The same approach can be used to develop better semiempirical models for other metal binding proteins (e.g. calcium and magnesium). If the simulations are long enough, they would also allow to study ligand exchange kinetics.
3. The optimized models reproduce gas-phase proton transfer energies in zinc hydrolases; the models could be used in a QM/MM representation to explore the multiple proton transfer pathways accessible to an enzyme, and the influence of neighboring residues on its reactivity.

7 References

[1] M.-C. Bellissent-Funel, A. Hassanali, M. Havenith, R. Henchman, P. Pohl, F. Sterpone, D. van der Spoel, Y. Xu, and A.E. Garcia, Chem. Rev. **116**, 7673 (2016).

[2] P. Ball, Chem. Rev. **108**, 74 (2008).

[3] A. Bertoluzza, C. Fagnano, M. Morelli, A. Tinti, and M. Tosi, J. Mol. Struct. **297**, 425 (1993).

[4] K. Amann-Winkel, M.-C. Bellissent-Funel, L.E. Bove, T. Loerting, A. Nilsson, A. Paciaroni, D. Schlesinger, and L. Skinner, Chem. Rev. **116**, 7570 (2016).

[5] P. Ball, Nature **452**, 291 (2008).

[6] F. Perakis, L. De Marco, A. Shalit, F. Tang, Z.R. Kann, T.D. Kühne, R. Torre, M. Bonn, and Y. Nagata, Chem. Rev. **116**, 7590 (2016).

[7] T. Fransson, Y. Harada, N. Kosugi, N.A. Besley, B. Winter, J.J. Rehr, L.G.M. Pettersson, and A. Nilsson, Chem. Rev. **116**, 7551 (2016).

[8] G.A. Cisneros, K.T. Wikfeldt, L. Ojamäe, J. Lu, Y. Xu, H. Torabifard, A.P. Bartók, G. Csányi, V. Molinero, and F. Paesani, Chem. Rev. **116**, 7501 (2016).

[9] K. Stokely, M.G. Mazza, H.E. Stanley, and G. Franzese, Proc. Natl. Acad. Sci. U. S. A. **107**, 1301 (2010).

[10] A. Nilsson and L.G.M. Pettersson, Nat. Commun. **6**, 8998 (2015).

[11] M.F. Chaplin, Biochem. Mol. Biol. Educ. **29**, 54 (2001).

[12] M.T. Sonoda, N.H. Moreira, L. Martinez, F.W. Favero, S.M. Vechi, L.R. Martins, and M.S. Skaf, Brazilian J. Phys. **34**, 3 (2004).

[13] E. Codorniu-Hernández and P.G. Kusalik, Proc. Natl. Acad. Sci. U. S. A. **110**, 13697 (2013).

[14] M.E. Tuckerman, D. Marx, and M. Parrinello, Nature **417**, 925 (2002).

[15] D. Marx, A. Chandra, and M.E. Tuckerman, Chem. Rev. **110**, 2174 (2010).

[16] C.A. Wraight, Biochim. Biophys. Acta - Bioenerg. **1757**, 886 (2006).

[17] Y. Peng, J.M.J. Swanson, S.-G. Kang, R. Zhou, and G.A. Voth, J. Phys. Chem. B **119**, 9212 (2014).

[18] W.L. Marshall and E.U. Franck, J. Phys. Chem. Ref. Data **10**, 295 (1981).

[19] A. V. Bandura and S.N. Lvov, J. Phys. Chem. Ref. Data **35**, 15 (2006).

[20] E. V. Anslyn and D.A. Dougherty, *Modern Physical Organic Chemistry* (University Science Books, Sausalito, California, 2006).

[21] H. Ishikita and K. Saito, J. R. Soc. Interface **11**, 20130518 (2014).

[22] C.J.T. de Grotthuss, Ann. Chim. **58**, 54 (1806).

[23] O. Markovitch and N. Agmon, J. Phys. Chem. A **111**, 2253 (2007).

[24] A. Hassanali, F. Giberti, J. Cuny, T.D. Kühne, and M. Parrinello, Proc. Natl. Acad. Sci. U. S. A. **110**, 13723 (2013).

[25] N. Agmon, H.J. Bakker, R.K. Campen, R.H. Henchman, P. Pohl, S. Roke, M. Thämer, and A. Hassanali, Chem. Rev. **116**, 7642 (2016).

[26] I. Persson, Pure Appl. Chem. **82**, 1901 (2010).

[27] L. Pauling, J. Am. Chem. Soc. **51**, 1010 (1929).

[28] J.K. Beattie, S.P. Best, B.W. Skelton, and A.H. White, J. Chem. Soc. Dalt. Trans. 2105 (1981).

[29] W. Rudolph, M.H. Brooker, and C.C. Pye, J. Phys. Chem. **99**, 3793 (1995).

[30] H.H. Loeffler and B.M. Rode, J. Chem. Phys. **117**, 110 (2002).

[31] G. Lamoureux and B. Roux, J. Phys. Chem. B **110**, 3308 (2006).

[32] T.W. Whitfield, S. Varma, E. Harder, G. Lamoureux, S.B. Rempe, and B. Roux, J. Chem. Theory Comput. **3**, 2068 (2007).

[33] F. Jalilehvand, D. Spångberg, P. Lindqvist-Reis, K. Hermansson, I. Persson, and M. Sandström, J. Am. Chem. Soc. **123**, 431 (2001).

[34] H. Ohtaki and T. Radnai, Chem. Rev. **93**, 1157 (1993).

[35] G. Johansson, Adv. Inorg. Chem. **39**, 159 (1992).

[36] N.A. Hewish, G.W. Neilson, and J.E. Enderby, Nature **297**, 138 (1982).

[37] G.J. Herdman and G.W. Neilson, J. Phys. Condens. Matter **4**, 649 (1992).

[38] K.L. Haas and K.J. Franz, Chem. Rev. **109**, 4921 (2009).

[39] K.A. McCall, C. Huang, and C.A. Fierke, J. Nutr. **130**, 1437S (2000).

[40] R. Wu, P. Hu, S. Wang, Z. Cao, and Y. Zhang, J. Chem. Theory Comput. **6**, 337 (2010).

[41] C.W. Bock, A.K. Katz, and J.P. Glusker, J. Am. Chem. Soc. **117**, 3754 (1995).

[42] J.P. Glusker, A.K. Katz, and C.W. Bock, Rigaku J. **16**, 8 (1999).

[43] W.N. Lipscomb and N. Sträter, Chem. Rev. **96**, 2375 (1996).

[44] J.J.R. Frausto da Silva and R.J.P. Williams, *The Biological Chemistry of the Elements: The Inorganic Chemistry of Life*, 2nd ed. (Oxford University Press, Oxford, UK, 2001).

[45] G. Lamoureux, In preparation.

[46] G. Pearson, J. Am. Chem. Soc. **85**, 3533 (1963).

[47] R.G. Pearson, Coord. Chem. Rev. **100**, 403 (1990).

[48] J.A. Lemire, J.J. Harrison, and R.J. Turner, Nat. Rev. Microbiol. **11**, 371 (2013).

[49] R.J.P. Williams, R. Inst. Chem. Rev. **1**, 13 (1968).

[50] Editorial, Nat. Chem. Biol. **4**, 143 (2008).

[51] H.B. Gray, Proc. Natl. Acad. Sci. U. S. A. **100**, 3563 (2003).

[52] S.J. Lippard, Nat. Chem. Biol. **2**, 504 (2006).

[53] T. Morgan, J. Myers, and B.H. Teow, Aust. NZ J. Med. **14**, 458 (1984).

[54] A. Aydemir-Köksoy, J. Ankara Med. Sch. **24**, 73 (2002).

[55] P. Várnai and K. Zakrzewska, Nucleic Acids Res. **32**, 4269 (2004).

[56] G. Peers and N.M.N. Price, Nature **441**, 341 (2006).

[57] K. Kaur, R. Gupta, S.A. Saraf, and S.K. Saraf, Compr. Rev. Food Sci. Food Saf. **13**, 358 (2014).

[58] M. Stefanidou, C. Maravelias, A. Dona, and C. Spiliopoulou, Arch. Toxicol. **80**, 1 (2006).

[59] D.S. Auld, Biometals **14**, 271 (2001).

[60] A.I. Anzellotti and N.P. Farrell, Chem. Soc. Rev. **37**, 1629 (2008).

[61] J. Blumberger, G. Lamoureux, and M.L. Klein, J. Chem. Theory Comput. **3**, 1837 (2007).

[62] V. Pelmenschikov, M.R.A. Blomberg, and P.E.M. Siegbahn, J. Biol. Inorg. Chem. **7**, 284 (2002).

[63] T. Szarvas, F. vom Dorp, S. Ergün, and H. Rübben, Nat. Rev. Urol. **8**, 241 (2011).

[64] W. Peng, J. Yan, Y. Wan, B. Wang, J. Tao, G. Yang, H. Pan, and J. Wang, J. Clin. Immunol. **32**, 1409 (2012).

[65] K. Maskos and W. Bode, Mol. Biotechnol. **25**, 241 (2003).

[66] W. Li, J. Wang, J. Zhang, and W. Wang, Curr. Opin. Struct. Biol. **30**, 25 (2015).

[67] H. Ode, M. Nakashima, S. Kitamura, W. Sugiura, and H. Sato, Front. Microbiol. **3**, 258 (2012).

[68] S.O. Nielsen, C.F. Lopez, G. Srinivas, and M.L. Klein, J. Phys. Condens. Matter **16**, R481 (2004).

[69] U.-H. Zaheer and J.D. Madura, *Frontiers in Computational Chemistry Volume 2: Computer Applications for Drug Design and Biomolecular Systems* (Elsevier, 2015).

[70] A. González-Lafont, T.N. Truong, and D.G. Truhlar, J. Phys. Chem. **95**, 4618 (1991).

[71] I. Rossi and D.G. Truhlar, Chem. Phys. Lett. **233**, 231 (1995).

[72] J.A. Pople, D.P. Santry, and G. a. Segal, J. Chem. Phys. **43**, S129 (1965).

[73] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, and J.J.P. Stewart, J. Am. Chem. Soc. **107**, 3902 (1985).

[74] W. Koch and M.C. Holthausen, *A Chemist's Guide to Density Functional Theory*, 2nd ed. (Wiley-VCH, New York, 2001).

[75] P. Fulde, *Electron Correlations in Molecules and Solids*, 3rd ed. (Springer, 2002).

[76] A.J. Cohen, P. Mori-Sánchez, and W. Yang, Chem. Rev. **112**, 289 (2012).

[77] T.J. Giese and D.M. York, J. Chem. Phys. **133**, 244107 (2010).

[78] P.J. Stephens, F.J. Devlin, C.F. Chabalowski, and M.J. Frisch, J. Phys. Chem. **98**, 11623 (1995).

[79] A.D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[80] C. Lee, W. Yang, and R.G. Parr, Phys. Rev. B **37**, 785 (1988).

[81] W. Kohn and L.J. Sham, Phys. Rev. **385**, A1133 (1965).

[82] J.A. Pople and D.L. Beveridge, *Approximate Molecular Orbital Theory* (McGraw-Hill Book Company, USA, 1970).

[83] C.C.J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).

[84] G.G. Hall, Proc. R. Soc. Ser. A **205**, 541 (1951).

[85] M.J.S. Dewar and W. Thiel, J. Am. Chem. Soc. **99**, 4899 (1977).

[86] J.J.P. Stewart, J. Comput. Chem. **10**, 209 (1989).

[87] E. Huckel, Z. Phys. **70**, 204 (1931).

[88] F. Feng, H. Wang, W. Fang, and J.-G. Yu, J. Theor. Comput. Chem. **8**, 691 (2009).

[89] M.I. Bernal-Uruchurtu, M.T.C. Martins-Costa, C. Millot, and M.F. Ruiz-López, J. Comput. Chem. **21**, 572 (2000).

[90] M.I. Bernal-Uruchurtu and M.F. Ruiz-López, Chem. Phys. Lett. **330**, 118 (2000).

[91] M.P. Repasky, J. Chandrasekhar, and W.L. Jorgensen, J. Comput. Chem. **23**, 1601 (2002).

[92] W. Thiel and A.A. Voityuk, J. Phys. Chem. **100**, 616 (1996).

[93] P. Imhof, F. Noe, S. Fischer, and J.C. Smith, J. Chem. Theory Comput. **2**, 1050 (2006).

[94] P. Charbonneau, Astrophys J Supp **101**, 309 (1995).

[95] L. Verlet, Phys. Rev. **159**, 98 (1967).

[96] B. Guillot, J. Chem. Phys. **95**, 1543 (1991).

[97] R. Ramírez, T. López-Ciudad, P. Kumar P, and D. Marx, J. Chem. Phys. **121**, 3973 (2004).

[98] R.G. Gordon, J. Chem. Phys. **43**, 1307 (1965).

[99] D.A. McQuarrie, *Statistical Mechanics* (Harper-Collins Publisher, New York, 1976).

[100] J.E. Davies, N.L. Doltsinis, A.J. Kirby, C.D. Roussev, and M. Sprik, J. Am. Chem. Soc. **124**, 6594 (2002).

[101] L. Bernasconi, E.J. Baerends, and M. Sprik, J. Phys. Chem. B **110**, 11444 (2006).

[102] V.R.I. Kaila and G. Hummer, Phys. Chem. Chem. Phys. **13**, 13207 (2011).

[103] K. Khistyaev, A. Golan, K.B. Bravaya, N. Orms, A.I. Krylov, and M. Ahmed, J. Phys. Chem. A **117**, 6789 (2013).

[104] T. Loerting and K.R. Liedl, J. Phys. Chem. A **105**, 5137 (2001).

[105] E. Vöhringer-Martinez and A. Toro-Labbé, J. Comput. Chem. **31**, 2642 (2010).

[106] O.F. Mohammed, D. Pines, J. Dreyer, E. Pines, and E.T.J. Nibbering, Science **310**, 83 (2005).

[107] K. Laasonen, M. Sprik, M. Parrinello, and R. Car, J. Chem. Phys. **99**, 9080 (1993).

[108] H.-S. Lee and M.E. Tuckerman, J. Chem. Phys. **126**, 164501 (2007).

[109] M. Del Ben, M. Schönherr, J. Hutter, and J. VandeVondele, J. Phys. Chem. Lett. **4**, 3753 (2013).

[110] R. Bukowski, K. Szalewicz, G.C. Groenenboom, and A. Van Der Avoird, J. Chem. Phys. **128**, 94314 (2008).

[111] R. Bukowski, K. Szalewicz, G.C. Groenenboom, and A. Van Der Avoird, Science **315**, 1249 (2007).

[112] T. Clark, J. Mol. Struct. THEOCHEM **530**, 1 (2000).

[113] J.W. Ponder and D.A. Case, Adv. Protein Chem. **66**, 27 (2003).

[114] C.M. Baker, Wiley Interdiscip. Rev. Comput. Mol. Sci. **5**, 241 (2015).

[115] J.A. Lemkul, J. Huang, B. Roux, and A.D. MacKerell, Chem. Rev. **116**, 4983 (2016).

[116] K.I. Ramachandran, G. Deepa, and N. K., *Computational Chemistry and Molecular Modeling: Principles and Applications* (Springer, Berlin, 2008).

[117] W. Thiel, Wiley Interdiscip. Rev. Comput. Mol. Sci. **4**, 145 (2014).

[118] A.S. Christensen, T. Kubar, Q. Cui, and M. Elstner, Chem. Rev. **116**, 5301 (2016).

[119] J.J.P. Stewart, J. Mol. Model. **13**, 1173 (2007).

[120] S. Wang, L. MacKay, and G. Lamoureux, J. Chem. Theory Comput. **10**, 2881 (2014).

[121] T.J. Giese, E.C. Sherer, C.J. Cramer, and D.M. York, J. Chem. Theory Comput. **1**, 1275 (2005).

[122] W. Harb, M.I. Bernal-Uruchurtu, and M.F. Ruiz-Lopez, Theor. Chem. Acc. **112**, 204 (2004).

[123] S. Liang and A.E. Roitberg, J. Chem. Theory Comput. **9**, 4470 (2013).

[124] X. Wu, W. Thiel, S. Pezeshki, and H. Lin, J. Chem. Theory Comput. **9**, 2672 (2013).

[125] M. Korth and W. Thiel, J. Chem. Theory Comput. **7**, 2929 (2011).

[126] R. Villar, M.J. Gil, J.I. García, and V. Martívez-Merino, J. Comput. Chem. **26**, 1347 (2005).

[127] K. Nam, Q. Cui, J. Gao, and D.M. York, J. Chem. Theory Comput. **3**, 486 (2007).

[128] J.J.P. Stewart, J. Mol. Model. **15**, 765 (2009).

[129] Z. Bikadi and E. Hazai, J. Cheminform. **1**, 1 (2009).

[130] G.B. Rocha, R.O. Freire, A.M. Simas, and J.J.P. Stewart, J. Comput. Chem. **27**, 1101 (2006).

[131] M. Bräuer, M. Kunert, E. Dinjus, M. Klußmann, M. Döring, H. Görls, and E. Anders, J. Mol. Struct. THEOCHEM **505**, 289 (2000).

[132] J.P. McNamara and I.H. Hillier, Phys. Chem. Chem. Phys. **9**, 2362 (2007).

[133] E.N. Brothers, D. Suárez, D.W. Deerfield, and K.M. Merz Jr., J. Comput. Chem. **25**, 1677 (2004).

[134] N.T. Anh, G. Frison, A. Solladié-Cavallo, and P. Metzner, Tetrahedron **54**, 12841 (1998).

[135] M.R. Silva-Junior and W. Thiel, J. Chem. Theory Comput. **6**, 1546 (2010).

[136] L. Xu, W. Cai, and X. Shao, J. Mol. Struct. THEOCHEM **817**, 35 (2007).

[137] J.J. Dannenberg, J. Mol. Struct. THEOCHEM **401**, 279 (1997).

[138] L.H. Coudert and J.T. Hougen, J. Mol. Spectrosc. **139**, 259 (1990).

[139] J.A. Odutola and T.R. Dyke, J. Chem. Phys. **72**, 5062 (1980).

[140] K.S. Kim, B.J. Mhin, U.-S. Choi, and K. Lee, J. Chem. Phys. **97**, 6649 (1992).

[141] W. Klopper, J.G.C.M. van Duijneveldt-van de Rijdt, and F.B. van Duijneveldt, Phys. Chem. Chem. Phys. **2**, 2227 (2000).

[142] W. Herndon and T. Radhakrishnan, Chem. Phys. Lett. **148**, 492 (1988).

[143] I.H. Williams, J. Am. Chem. Soc. **109**, 6299 (1987).

[144] A.A. Bliznyuk and A.A. Voityuk, J. Mol. Struct. THEOCHEM **164**, 343 (1988).

[145] G. Monard, M.I. Bernal-Uruchurtu, A. van der Vaart, K.M.J. Merz, and M.F. Ruiz-López, J. Phys. Chem. A **109**, 3425 (2005).

[146] G. Murdachaew, C.J. Mundy, G.K. Schenter, T. Laino, and J. Hutter, J. Phys. Chem. A **115**, 6046 (2011).

[147] M. Welborn, J. Chen, L.-P. Wang, and T. Van Voorhis, J. Comput. Chem. **36**, 934 (2015).

[148] Y. Lin, A. Wynveen, J.W. Halley, L.A. Curtiss, and P.C. Redfern, J. Chem. Phys. **136**, 174507 (2012).

[149] V.S. Bryantsev, M.S. Diallo, A.C.T. Van Duin, and W.A. Goddard, J. Chem. Theory Comput. **5**, 1016 (2009).

[150] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J. V Ortiz, J. Cioslowski, and D.J. Fox, Gaussian Inc Wallingford CT **34**, Wallingford CT (2009).

[151] J.J.P. Stewart, J. Comput. Aided. Mol. Des. **4**, 1 (1990).

[152] CP2K, www.cp2k.org.

[153] G. Klopman, J. Am. Chem. Soc. **86**, 4550 (1964).

[154] M.J.S. Dewar and N.L. Hojvat, J. Chem. Phys. **34**, 1232 (1961).

[155] M.J.S. Dewar and N.L. Hojvat, Proc. R. Soc. A Math. Phys. Eng. Sci. **264**, 431 (1961).

[156] M.J.S. Dewar and N.L. Sabelli, J. Phys. Chem. **66**, 2310 (1962).

[157] K. Ohno, Theor. Chim. Acta **2**, 219 (1964).

[158]PT structures are two from B2 in tetra, one from C5 in penta, and two from D8 in hexa of Figure 2 of ref. [120].

[159] A.K. Soper, Chem. Phys. **258**, 121 (2000).

[160] T. Todorova, A.P. Seitsonen, J. Hutter, I.-F.W. Kuo, and C.J. Mundy, J. Phys. Chem. B **110**, 3685 (2006).

[161] A.K. Soper, F. Bruni, and M.A. Ricci, J. Chem. Phys. **106**, 247 (1997).

[162] K.N. Marsh, *Recommended Reference Materials for the Realization of Physicochemical Properties* (Blackwell, Oxford, 1987).

[163] T. Fox and P.A. Kollman, J. Phys. Chem. B **102**, 8070 (1998).

[164] L.X. Dang, J. Phys. Chem. B **102**, 620 (1998).

[165] S.A. Clough, Y. Beers, G.P. Klein, and L.S. Rothman, J. Chem. Phys. **59**, 2254 (1973).

[166] T.R. Dyke, K.M. Mack, and J.S. Muenter, J. Chem. Phys. **66**, 498 (1977).

[167] Y.S. Badyal, M.L. Saboungi, D.L. Price, S.D. Shastri, D.R. Haeffner, and A.K. Soper, J. Chem. Phys. **112**, 9206 (2000).

[168] I.C. Yeh and G. Hummer, J. Phys. Chem. B **108**, 15873 (2004).

[169] R. Mills, J. Phys. Chem. **77**, 685 (1973).

[170] D.P. Fernandez, Y. Mulev, A.R.H. Goodwin, and J.M.H. Levelt Sengers, J. Phys. Chem. Ref. Data **24**, 33 (1995).

[171] D. Laage and J.T. Hynes, Science **311**, 832 (2006).

[172] M. Neumann and O. Steinhauser, Chem. Phys. Lett. **106**, 563 (1984).

[173] M. Sprik, J. Chem. Phys. **95**, 6762 (1991).

[174] N. Yoshii, S. Miura, and S. Okazaki, Chem. Phys. Lett. **345**, 195 (2001).

[175] M. Neumann, O. Steinhauser, and G.S. Pawley, Mol. Phys. **52**, 97 (1984).

[176] S.W. De Leeuw, J.W. Perram, and E.R. Smith, Annu. Rev. Phys. Chem. **37**, 245 (1986).

[177] S.W. Rick, S.J. Stuart, and B.J. Berne, J. Chem. Phys. **101**, 6141 (1994).

[178] S.J. Rosenthal, R. Jimenez, G.R. Fleming, P. V Kumar, and M. Maroncelli, J. Mol. Liq. **60**, 25 (1994).

[179] A.Y. Zasetsky, Phys. Rev. Lett. **107**, 117601 (2011).

[180] U. Kaatze, J. Chem. Eng. Data **34**, 371 (1989).

[181] J. Kim, J.Y. Lee, S. Lee, B.J. Mhin, and K.S. Kim, J. Chem. Phys. **102**, 310 (1995).

[182] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II* (Springer, Berlin, 1991).

[183] P. Schofield, Phys. Rev. Lett. **4**, 239 (1960).

[184] P.L. Silvestrelli, M. Bernasconi, and M. Parrinello, Chem. Phys. Lett. **277**, 478 (1997).

[185] S.T. Roberts, K. Ramasesha, and A. Tokmakoff, Acc. Chem. Res. **42**, 1239 (2009).

[186] C.P. Lawrence and J.L. Skinner, J. Chem. Phys. **118**, 264 (2003).

[187] R. Rey, K.B. Møller, and J.T. Hynes, J. Phys. Chem. A **106**, 11993 (2002).

[188] J.E. Bertie and Z. Lan, Appl. Spectrosc. **50**, 1047 (1996).

[189] S.Y. Venyaminov and F.G. Prendergast, Anal. Biochem. **248**, 234 (1997).

[190] J. Anastassopoulou and T. Theophanides, Bioinorg. Chem. **459**, 209 (1995).

[191] W.J. Song, P.A. Sontz, X.I. Ambroggio, and F.A. Tezcan, Annu. Rev. Biophys. **43**, 409 (2014).

[192] M. Laitaoja, J. Valjakka, and J. Jänis, Inorg. Chem. **52**, 10983 (2013).

[193] N. Pace and E. Weerapana, Biomolecules **4**, 419 (2014).

[194] L. Helm and A.E. Merbach, Coord. Chem. Rev. **187**, 151 (1999).

[195] B.L. Vallee and D.S. Auld, Proc. Natl. Acad. Sci. U. S. A. **87**, 220 (1990).

[196] Y.-M. Lee, Y.-F. Lin, and C. Lim, J. Chinese Chem. Soc. **61**, 142 (2014).

[197] T. Kochańczyk, A. Drozd, and A. Krężel, Metallomics **7**, 244 (2015).

[198] J. Zhang, W. Yang, J.-P. Piquemal, and P. Ren, J. Chem. Theory Comput. **8**, 1314 (2012).

[199] C.W. Bock, A.K. Katz, G.D. Markham, and J.P. Glusker, J. Am. Chem. Soc. **121**, 7360 (1999).

[200] K. Patel, A. Kumar, and S. Durani, Biochim. Biophys. Acta - Proteins Proteomics **1774**, 1247 (2007).

[201] B. De Courcy, N. Gresh, and J.P. Piquemal, Interdiscip. Sci. Comput. Life Sci. **1**, 55 (2009).

[202] N. Gresht and D.R. Garmer, J. Comput. Chem. **17**, 1481 (1996).

[203] V.M. Rayón, H. Valdés, N. Díaz, and D. Suárez, J. Chem. Theory Comput. **4**, 243 (2008).

[204] E.A. Amin and D.G. Truhlar, J. Chem. Theory Comput. **4**, 75 (2008).

[205] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).

[206] S.J. Weiner, P.A. Kollman, D.T. Nguyen, and D.A. Case, J. Comp. Chem. **7**, 230 (1986).

[207] W.L. Jorgensen and J. Tirado-Rives, J. Am. Chem. Soc. **110**, 1657 (1988).

[208] W.L. Jorgensen, D.S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).

[209] J.C. Wu, J.-P. Piquemal, R. Chaudret, P. Reinhardt, and P. Ren, J. Chem. Theory Comput. **6**, 2059 (2010).

[210] R. Wu, Z. Lu, Z. Cao, and Y. Zhang, J. Chem. Theory Comput. **7**, 433 (2011).

[211] S. Riahi, B. Roux, and C.N. Rowley, Can. J. Chem. **99**, 1 (2013).

[212] T. Zhu, X. Xiao, C. Ji, and J.Z.H. Zhang, J. Chem. Theory Comput. **9**, 1788 (2013).

[213] M.B. Peters, Y. Yang, B. Wang, L. Füsti-Molnár, M.N. Weaver, and K.M. Merz Jr., J. Chem. Theory Comput. **6**, 2935 (2010).

[214] F. Duarte, P. Bauer, A. Barrozo, B.A. Amrein, M. Purg, J. Åqvist, and S.C.L. Kamerlin, J. Phys. Chem. B **118**, 4351 (2014).

[215] G. Frison and G. Ohanessian, J. Comput. Chem. **29**, 416 (2008).

[216] X. Lopez and D.M. York, Theor. Chem. Accounts Theory, Comput. Model. (Theoretica Chim. Acta) **109**, 149 (2003).

[217] K. Govender, J. Gao, and K.J. Naidoo, J. Chem. Theory Comput. **10**, 4694 (2014).

[218] I. Tejero, A. González-Lafont, and J.M. Lluch, J. Comput. Chem. **28**, 997 (2007).

[219] J.P. McNamara, M. Sundararajan, I.H. Hillier, J. Ge, A. Campbell, and C. Morgado, J. Comput. Chem. **27**, 1307 (2006).

[220] B.K. Sharma, S. Wang, and G. Lamoureux, In preparation.

[221] D.R. Stull and H. Prophet, *JANAF Thermochemical Tables*, 2nd ed. (National Bureau of Standards:Washington DC, 1971).

[222] M.J.S. Dewar and K.M. Merz, Organometallics **7**, 522 (1988).

[223] M.J.S. Dewar and Y. Yuan, Inorg. Chem. **29**, 3881 (1990).

[224] N. Rega, G. Brancato, A. Petrone, P. Caruso, and V. Barone, J. Chem. Phys. **134**, 74504 (2011).

[225] A. Kuzmin, S. Obst, and J. Purans, J. Phys. Condens. Matter **9**, 10065 (1997).

[226] E. Cauët, S. Bogatko, J.H. Weare, J.L. Fulton, G.K. Schenter, and E.J. Bylaska, J. Chem. Phys. **132**, 194502 (2010).

[227] W. Maret and Y. Li, Chem. Rev. **109**, 4682 (2009).

[228] M.D. Liptak and G.C. Shields, J. Am. Chem. Soc. **123**, 7314 (2001).

[229] M. Sprik, Chem. Phys. **258**, 139 (2000).

[230] D.D. Perrin, J. Chem. Soc. 4500 (1962).

[231] S.J. Hawkes, J. Chem. Educ. **73**, 516 (1996).

[232] M.I. Page, *Amino Acids, Peptides and Proteins in Organic Chemistry* (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2010).

[233] P.A. Frey and A.D. Hegeman, *Enzymatic Reaction Mechanisms* (Oxford University Press, Oxford, UK, 2007).

[234] A. Streitwieser and C.H. Heathcock, *Introduction to Organic Chemistry* (Macmillian Publishing Co. Inc., New York, 1985).

[235] M.N. Hughes, M.N. Centelles, and K.P. Moore, Free Radic. Biol. Med. **47**, 1346 (2009).

[236] L. Li, A. Hsu, and P.K. Moore, Pharmacol. Ther. **123**, 386 (2009).

[237] R.O. Beauchamp Jr., J.S. Bus, J.A. Popp, C.J. Boreiko, and D.A. Andjelkovich, Crit. Rev. Toxicol. **13**, 25 (1984).

[238] D.R. Lide, *CRC Handbook of Chemistry and Physics*, 84th ed. (CRC Press, Boca Raton, FL, 2004).

[239] E. Łowicka and J. Bełtowski, Pharmacol. Reports **59**, 4 (2007).

[240] B. Akabayov, S.J. Lee, S.R. Akabayov, S. Rekhi, B. Zhu, and C.C. Richardson, Biochemistry **48**, 1763 (2009).

[241] T. Dudev and C. Lim, J. Am. Chem. Soc. **124**, 6759 (2002).

[242] B.L. O'Dell, Nutr. Rev. **50**, 48 (1992).

[243] W. Maret, J. Trace Elem. Med. Biol. **19**, 7 (2005).

[244] S. Iyer, R. Visse, H. Nagase, and K.R. Acharya, J. Mol. Biol. **362**, 78 (2006).

[245] D. Riccardi, S. Yang, and Q. Cui, Biochim. Biophys. Acta - Proteins Proteomics **1804**, 342 (2010).

[246] P. Tao, J.F. Fisher, Q. Shi, T. Vreven, S. Mobashery, and H.B. Schlegel, Biochemistry **48**, 9839 (2009).

[247] N. Díaz, D. Suarez, and T.L. Sordo, J. Phys. Chem. B **110**, 24222 (2006).

[248] W.B. De Almeida, H.F. Dos Santos, W.R. Rocha, and M.C. Zerner, J. Chem. Soc., Dalt. Trans. **1**, 2531 (1998).

[249] I. Adejoro and O. Oyeneyin, Am. J. Sci. Ind. Res. **4**, 111 (2013).

[250] Q. Cui, H. Guo, and M. Karplus, J. Chem. Phys. **117**, 5617 (2002).

[251] S. Antonczak, G. Monard, M.F. Ruiz-López, and J.-L. Rivail, J. Am. Chem. Soc. **120**, 8825 (1998).

[252] B.K. Sharma and G. Lamoureux, In preparation.

[253] V. Navrátil, V. Klusák, and L. Rulíšek, Chem. - A Eur. J. **19**, 16634 (2013).

[254] R.E. Hubbard and M. Kamran Haider, Encycl. Life Sci. John Wiley Sons, Ltd. 1 (2001).

[255] V.M. Krishnamurthy, G.K. Kaufman, A.R. Urbach, I. Gitlin, K.L. Gudiksen, D.B. Weibel, and G.M. Whitesides, Chem. Rev. **108**, 946 (2008).

[256] D.P. Martin, Z.S. Hann, and S.M. Cohen, Inorg. Chem. **52**, 12207 (2013).

[257] S. Jo, T. Kim, V.G. Iyer, and I. Wonpil, J. Comput. Chem. **29**, 1859 (2008).

[258] P. Goyal, H.-J. Qian, S. Irle, X. Lu, D. Roston, T. Mori, M. Elstner, and Q. Cui, J. Phys. Chem. B **118**, 11007 (2014).

[259] P. Goyal, M. Elstner, and Q. Cui, J. Phys. Chem. B **115**, 6790 (2011).

[260] S. Keskin and S. Kızılel, Ind. Eng. Chem. Res. **50**, 1799 (2011).

[261] S. Rojas, F.J. Carmona, C.R. Maldonado, P. Horcajada, T. Hidalgo, C. Serre, J.A.R. Navarro, and E. Barea, Inorg. Chem. **55**, 2650 (2016).

[262] J.R. Errington and P.G. Debenedetti, Nature **409**, 318 (2001).