

COMPUTATIONAL STUDIES OF THE STRUCTURE AND FUNCTION OF
METALLOENZYMES AND THE PERFORMANCE OF DENSITY FUNCTIONAL
METHODS

By

BRYAN T. OP'T HOLT

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2006

For Jenny Kay

ACKNOWLEDGMENTS

For the motivation to complete this dissertation and to focus on my research even when I thought it would be too hard, Jenny should get all the credit. For the ideas and suggestions that kindled my interest in computational and biological chemistry when it started to wane, I recognize Kennie. The desire to complete my educational career with a Ph.D. and the knowledge of the value of a strong education were instilled in me by my parents, and I owe them my best. As for life in the Merz group, I will always be thankful for the contributions of Ken Ayers, Ed Brothers, Guanglei Cui, and Kevin Riley to my research.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	ix
CHAPTER	
1 INTRODUCTION	1
2 QM AND MM METHODS OF THE STUDY OF METALLOPROTEINS	6
Metal Binding Site Studies in Gaussian 03	6
Molecular Mechanics of HAH1	10
The AMBER Force Field	10
LeAP and <i>sander</i>	12
Free Energy Simulations of HAH1	17
Thermodynamic Integration	20
Free Energy Perturbation	23
Potential of Mean Force	24
3 COMPUTATIONAL STUDIES OF THE CU(I) METALLOCHAPERONE HAH1	26
Cu(I) Biochemistry and the HAH1 System	26
Cu Metallochaperones	26
Cu pathways within the Cell	29
Copper Homeostasis	31
Quantum Chemical Characterization of the Cu(I) Binding Site from HAH1	33
Creation of a MM Force Field for the Cu(I) Binding Site of HAH1	36
The use of the HAH1 dimer as a model for the HAH1-MNK4 heterodimer	36
Creation of parameter files for Cu(I)-bound HAH1	37
Free Energy Calculations of the Cu(I)-Bound HAH1 Dimer	47
Quantum Energy Calculations on the Small Cu(I) Thiolate Cluster Models	48
Free Energy Calculations on HAH1	50
Conclusions	64

4	ELECTRONIC STRUCTURE OF THE ACTIVE SITE OF AMINOPEPTIDASE FROM <i>Aeromonas proteolytica</i>	66
	AAP Introduction.....	66
	Effects of 1 st -Shell Mutations	72
	Conclusions.....	81
5	SURVEY OF DENSITY FUNCTIONAL THEORY METHODS	83
	Introduction.....	83
	Methods	84
	Computational Methods.....	99
	Results by Property	117
	Bond Lengths.....	117
	Bond Angles	120
	Ground State Vibrational Frequencies	125
	Ionization Potential.....	129
	Electron Affinity.....	134
	Heat of Formation.....	139
	Hydrogen Bonding Interaction Energy	146
	Conformational Energies.....	150
	Reaction Barrier Heights for Small Systems with Non-singlet Transition States.....	158
	Reaction Barrier Heights for Organic Molecules with Singlet Transition States.....	159
	Conclusions.....	164
	General Summary of the Survey of DFT Methods	170
6	CONCLUSIONS	178
	APPENDIX	180
	LIST OF REFERENCES.....	182
	BIOGRAPHICAL SKETCH	193

LIST OF TABLES

<u>Table</u>	<u>page</u>
2-1 The DZpdf basis function used for Cu(I) in the DFT calculations of the cluster models.	9
2-2 Windows and weighting factors for a 12-window TI calculation in <i>sander</i>	22
2-3 λ values for a 12-window TI calculation in <i>sander</i>	22
3-1 Geometry parameters of the DFT-optimized multi-coordinate models shown in Figure 3-2.	36
3-2 Atom type, atomic mass, van der Waals radii, and van der Waals well-depths for Cu(I) and Cu(I)-bound S in HAH1.	41
3-3 Bond lengths, bond angles, and associated force constants for the HAH1 Cu(I) binding site.	41
3-4 CYM and Cu(I) RESP charges used for the HAH1 Cu(I) binding site.	41
3-5 Comparison of the HAH1 active site between the four-coordinate model, the solvated, equilibrated protein, and the X-ray crystal structure of the Cu(I)-bound protein (1FEE).	43
3-6 Summary of rms deviation, rms flexibility, and radius of gyration for the five solvated HAH1 protein models and key active site residues.	43
3-7 Reaction energies and energies of solvation for the model systems.	50
3-8 Atoms, atom types, atomic masses and van der Waals parameters used in TI simulations of Cu(I)-bound HAH1.	56
3-9 Bond length parameters for the reactions used in TI calculations of Cu(I)-bound HAH1.	57
3-10 Bond angle parameters for the reactions used in TI calculations of Cu(I)-bound HAH1.	58
3-11 RESP charges used for TI calculations on Cu(I)-bound HAH1.	59
3-12 Free energy changes for TI calculations on the reactions shown in Figure 3-9.	63

3-13 The free energy difference of changing the coordination environment of Cu(I) in HAH1.	63
4-1 Electrons in the side chains of Asp117 and Asp179 are equally delocalized over the carboxylic acid region, while Glu151 and Glu152 side chains contain one C-O bond with more electron density than the other.	76
4-2 Several distances are shown for Zn-Zn and Zn-O interactions for the structure shown below in Figures 4-6 and 4-7.	77
5-1 The thirty-seven density functional method and two wave function methods tested in this survey with appropriate references.....	90
5-2 Basis sets employed in the survey of DFT methods.....	97
5-3 Valence shell polarization functions incorporated into the correlation-consistent basis sets of Dunning.....	98
5-4 The bond lengths and bond angles test set.	102
5-5 The test set for ground state vibrational frequency.	104
5-6 The ionization potential test set.	106
5-7 The electron affinity test set.	107
5-8 The heat of formation test set, singlets only.....	108
5-9 The heat of formation test set, radicals only.....	111
5-10 The hydrogen bonding test set.....	113
5-11 The conformational energy test set.....	114
5-12 The small system radical transition state reaction barrier test set.	115
5-13 The organic molecule singlet transition state reaction barrier test set.....	116
5-14 Bond lengths, bond angles, and vibrational frequencies for HF, MP2, and LSDA methods.	119
5-15 Average unsigned ionization potential errors for HF, MP2, LSDA, and B3P86 methods.	134
5-16 Average unsigned electron affinity errors for HF, MP2, LSDA, and B3P86 methods.	139
5-17 Average unsigned HOF errors for the HF, MP2, and LSDA methods.....	143

5-18 Average unsigned hydrogen bond interaction errors for HF, MP2, and LSDA methods.	150
5-19 Average unsigned conformational energy errors for HF, MP2, and LSDA methods.	155
5-20 Average unsigned errors for the non-singlet transition state reaction test set for HF, MP2, and LSDA methods.	159
5-21 Average unsigned errors for the singlet transition state large reaction test set for HF, MP2, and LSDA methods.	164
5-22 Rankings of functional/basis set combinations for properties considered in this work.	171
5-23 Average functional rankings and standard deviations for the top fifteen functionals along with 6-31+G* and aug-cc-pVDZ basis sets.	175
5-24 Performances of the fifteen highest-ranking functionals paired with the 6-31G* basis.	176
5-25 Performances of the fifteen highest ranked density functional methods paired with the aug-cc-pVDZ basis set.	177

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
3-1 The proposed mechanism for Cu(I) transfer from HAH1 to the fourth domain of the Wilson's disease protein. Cym indicates a negatively charged Cys residue.	34
3-2 Gas-phase optimized structures of the multi-coordinate models of the HAH1 Cu(I) binding site. Cu(I) is in green, S is yellow, and C is gray. The top figure shows the two-coordinate model, followed by the three-coordinate model and the four-coordinate model at the bottom.	35
3-3 The 1.80 Å crystal structure for Cu(I)-bound HAH1. PDB ID 1FEE. ¹	40
3-4 Root-mean-squared deviations between the five solvated HAH1 models and the Cu(I)-bound HAH1 crystal structure as a function of time.	45
3-5 rmsd between the active site loop regions of the five solvated HAH1 models and the Cu(I)-bound HAH1 crystal structure as a function of time.	46
3-6 The isodesmic reaction and solvation of three-coordinate Cu(I) to four-coordinate in the model system.	49
3-7 The relative energies of the species in the isodesmic reactions of the model systems in the gas phase (top) and in implicit solvent.	50
3-8 A scheme for Cu(I) transfer.	52
3-9 Two separate TI simulations were performed to compare two different Cu(I) transfer pathways from HAH1 to MNK4. The two reactions shown have different starting points, but the same endpoint. The top reaction is referred to as "Reaction 1" and the bottom is "Reaction 2" in the discussion that follows.	53
3-10 FEP vdW correction to TI on HAH1. Evaluate trajectory of structure with vdW exclusions removed with the Hamiltonian with vdW exclusions intact.	60
3-11 Bond length correction to FEP calculations by PMF: contract Cu(I)-S bond length from ~2.8 Å to ~2.0 Å by PMF analysis of twenty-three windows.	61
3-12 PMF curve of solvated HAH1 showing minimum energy Cu(I)-S _{12B} bond length near 2.1 Å for the bonding of Cys 12B to Cu(I).	62

3-13	PMF curve for the binding of Cys 15B to Cu(I) in solvated HAH1, showing a minimum energy bond-length of just over 2.1 Å for the Cu(I)-S _{15B} bond.....	62
3-14	The free energy difference by thermodynamic integration between the different three-coordinate Cu(I)-bound HAH1 dimers.	63
4-1	AAP active site inhibited by Tris (left) and BuBA . Investigation of the X-ray structures of these complexes shed light on substrate conformation and a potential mechanism for peptide hydrolysis in AAP. PDB ID 1LOK, 1CP6	67
4-2	In fluoride inhibition studies of AAP, it was shown that a F ⁻ ion displaces a terminal hydroxide group, deactivating the enzyme.	68
4-3	A proposed mechanism for AAP peptide hydrolysis showing proton transfer to Glu151, formation of a terminal hydroxyl- group, a gem-diolate intermediate, donation of a proton back to the leaving amino group, and reformation of the water/hydroxide bridge. Adapted from Petsko. ⁵⁹	70
4-4	The general model for the QM work is the AAP active site from PDB structure 1AMP, the 1.8 Å resolution structure elucidated by Chevrier and Schalk. ⁶⁷ Asp 117 is below the two Zn ions, with Zn ₂ on the left and Zn ₁ on the right. The residue at the top of the active site is Glu 151. Zn ₂ is bound to His 97 and Asp 179, and Zn ₁ is complexed with His 256 and Glu 152.....	74
4-5	B3LYP/6-31G* optimized geometries of two models of the AAP active site. Asp 117 is shown in the upper-right of both pictures, binding each Zn. Zn ₁ is the ion on the left and Zn ₂ on the right in each structure. The structure on the left is from an originally water-bridged structure and Glu151 has gained a H, while the structure on the right started with a OH ⁻ bridge and Asp 179 gains a H from a crystallographic water.	78
4-6	Starting structures (left) and B3LYP/6-31G* optimized geometries for different Zn-Zn bridging species within the active site of AAP. a) a water bridge b) a hydroxyl- bridge c) an oxyl- bridge.	79
4-7	Initial structures (left) and B3LYP/6-31G* optimized geometries for models with Glu151 protonated. The starting structures vary only in the protonation state of the bridging group. Structure a) contains an O ²⁻ bridge and the Zn ions in c) are bridged by a hydroxide group.	80
5-1	Average unsigned bond length errors (in Å) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Pople type basis sets.	121
5-2	Average unsigned bond length errors (in Å) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Dunning type basis sets.	122
5-3	Average unsigned bond angle errors (in degrees) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Pople type basis sets.	126

5-4 Average unsigned bond angle errors (in degrees) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Dunning type basis sets.	127
5-5 Average unsigned vibrational frequency errors (in cm^{-1}) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Pople type basis sets.	130
5-6 Average unsigned vibrational frequency errors (in cm^{-1}) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Dunning type basis sets.	131
5-7 Average unsigned ionization potential errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals with Pople type basis sets.	135
5-8 Average unsigned ionization potential errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals with Dunning type basis sets.	136
5-9 Average unsigned electron affinity errors (kcal/mol) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals.	140
5-10 Average unsigned heat of formation errors (kcal/mol) for the five Pople-style basis sets employed in this study.	144
5-11 Average unsigned heat of formation errors (kcal/mol) for the Dunning-type basis functions used in this work.	145
5-12 Average unsigned hydrogen bond interaction energy errors (kcal/mol) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Pople type basis sets.	151
5-13 Average unsigned hydrogen bond interaction energy errors (kcal/mol) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Dunning type basis sets.	152
5-14 Average unsigned conformational energy errors (kcal/mol) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Pople-type basis sets.	156
5-15 Average unsigned conformational energy errors (kcal/mol) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Dunning-type basis sets.	157
5-16 Average unsigned barrier height energy errors (kcal/mol) for SRBH reactions along with the GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Pople type basis sets.	160

5-17 Average unsigned barrier height energy errors (kcal/mol) for SRBH reactions along with the GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Dunning-type basis sets.	161
5-18 Average unsigned barrier height energy errors (kcal/mol) for large singlet transition state reactions along with the GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Pople-type basis sets.	165
5-19 Average unsigned barrier height energy errors (kcal/mol) for large singlet transition state reactions along with the GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Dunning-type basis sets.	166

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

COMPUTATIONAL STUDIES OF THE STRUCTURE AND FUNCTION OF
METALLOENZYMES AND THE PERFORMANCE OF DENSITY FUNCTIONAL
METHODS

By

Bryan T. Op't Holt

December 2006

Chair: Kenneth M. Merz, Jr.
Major Department: Chemistry

Transition metals work within metalloproteins to catalyze an innumerable array of important reactions in the body. Experimental data on the intermolecular and intramolecular phenomena that govern these biochemical processes are often difficult to obtain with traditional methods of analytical and biological chemistry. However, computational methods have been developed to simulate macromolecular systems in order to understand the dynamics and energetics of protein structure and function.

The first half of this dissertation details the application of quantum mechanical and molecular mechanical methods to the metalloproteins HAH1 and aminopeptidase. The active site of the Cu(I)-binding protein HAH1 is initially characterized using QM methods to determine the geometrical and electrostatic parameters of the system. This information is subsequently used to create a molecular mechanics force field for the active site within the native holo-protein. A series of molecular dynamics simulations and free energy calculations are performed on the protein following this parameterization.

The energetics of Cu(I) transfer are discussed and a favorable mechanism for metal ion transport is proposed. The aminopeptidase system is investigated in a different manner as the effects of first-shell mutations are studied with respect to the electronic structure and coordination of the di-Zn(II) active site. *Ab initio* methods are employed to describe a series of native and mutant active sites.

The second part of this work describes a large-scale survey of the performance of density functional methods and basis sets at predicting a series of molecular properties and intermolecular interactions. As the computational resources available to the scientific community increase in speed while becoming more affordable, the study of large scale systems with DFT and *ab initio* methods will be more plausible. The intention of this study is to individually examine the accuracy of density functional methods, not only to identify which methods work the best for certain properties, but also to note where there may be room for improvement as more methods are developed.

CHAPTER 1 INTRODUCTION

As the scientific community continues to discover more about the roles of metalloenzymes in biological processes, it becomes necessary to investigate the structure and function of these proteins on numerous levels. Experimental methods have been used to study the structure and function of proteins for decades, and now computational methods are available that can be used to further investigate protein systems. The combination of experimental and computational research on biochemical systems enhances the ability to identify potential drug targets, determine catalytic mechanisms of protein function, and describe the energetics of chemistry occurring in the body on the macromolecular and atomic levels.

For example, X-ray crystallography and NMR spectroscopy visualize the structures of proteins. Spectroscopic methods such as EXAFS, XAS, and EPR can be used to elucidate conformational and electronic details of the metal binding sites of metalloenzymes in certain cases, and kinetics and binding affinity studies are routinely used to suggest mechanisms of protein function. However, each of these methods has limitations as the scale of the research approaches the atomic and sub-atomic levels.

Except for the most recent high-resolution structures, X-ray crystallography is not able to fully resolve the protonation states of residues, which is a key factor in protein function. The practical difficulty of the method is compounded by the delicate task of crystallizing a sample (only to have it destroyed by the experiment). Dynamically, X-ray structures reveal the protein's structure at low temperature and in a fixed crystalline

state.¹ On the other hand, NMR methods can sample many conformations of proteins in solution and at room temperature and provide high structural resolution.² Wet chemistry techniques such as assays, titrations, and electrophoresis supply valuable data on reaction rates, equilibrium constants, and binding affinity, but are time consuming. Moreover, the number of distinct systems that can be evaluated consecutively is limited by lab space and lab equipment. A trait that is missing from all of these experimental methods is the ability to predict or measure the energetics of the individual bonds and interactions that play key roles in protein structure and function.

The past twenty years has seen extensive advancements in computational methods and computational ability. From the semi-empirical methods of Dewar *et al.*³ to the newest density functional methods of the Scuseria and Truhlar groups,^{4,5} systems are being better-defined both energetically and geometrically. In addition, more complex systems are being investigated today than ever before thanks to massively parallel computer clusters that feature extensive available memory and the ability to perform trillions of operations per second. Furthermore, development of software packages such as AMBER 9.0⁶ and Gaussian 03⁷ allow researchers to carry out a variety of specific calculations to meet their needs.

While the ability to computationally predict or determine protein structure or sequence has not been developed, many of the systematic hurdles of the experimental methods mentioned above can be addressed by some form of presently available computational method. Reliable, accurate data on proteins and parts of proteins can be gathered using robust and well-validated computational techniques. Bonding and nonbonding interaction energies can be quantified, electrostatic potentials can be

calculated, strained systems can be energetically minimized and equilibrated, and the motion of large systems can be followed over time in dynamics simulations. Moreover, computational methods can be amended or parameterized for accuracy and new methods can be developed that scale less dramatically or are not heavily parameter-dependent.

This dissertation highlights the application of computational methods to a range of chemical systems from small molecules to solvated proteins. Together with experimental data, the application of computational techniques provides a complete picture of the world of biochemistry. The following chapters detail the investigation of metal ion transfer between proteins, the electronic structure of a di-metal active site, and the assessment of the performance of commonly used density functional methods.

Chapter 2 outlines the techniques used in the computational study of metalloproteins. Starting with the use of QM methods to describe the metal binding sites of the Cu(I) binding protein HAH1 and the di-Zn(II) site of the aminopeptidase from *A. proteolytica* (AAP), the process of defining the geometries and energetics of metal binding sites is explained. The QM data for the HAH1 system were used to construct a MM force field for the Cu(I) binding site. The force field was employed in a long timescale MD simulation to investigate the dynamics of Cu(I)-bound HAH1. Furthermore, the free energy of Cu(I) transfer from HAH1 to the fourth domain of the Menkes Disease Protein, MNK4 is investigated using other MM methods including thermodynamic integration, free energy perturbation, and potential of mean force calculations.

Chapter 3 discusses Cu(I) homeostasis, Cu(I) transport, and the HAH1 system. QM calculations using Gaussian 03⁷ were conducted on a model of the HAH1 active site

based on the crystal structure of the HAH1 dimer, and MM calculations were performed on the entire HAH1 dimer in gas-phase and explicit solvent with AMBER.⁶ The ultimate goal of the HAH1 study was to elucidate details of Cu(I) transfer from HAH1 to the Menkes disease protein MNK. This study involved the QM description of the HAH1 active site, the construction of an MM force field to describe the active site classically, and numerous MD-based calculations. Finally, the results of the study are discussed in terms of protein dynamics and the mechanistic implications of the research.

Chapter 4 summarizes the application of QM methods to the AAP system. Several questions regarding the active site were addressed by modeling variations of the active site of the protein. The first goal was to identify the molecular bridge between the two Zn(II) ions in the active site. The coordination state of each ion was investigated, and the contribution of each ligating residue was studied.

The fifth chapter details the large-scale survey of the performance of commonly used density functional methods and basis functions. The accuracy of thirty-seven density functionals and two wave function methods is evaluated for nine molecular and intermolecular properties. These properties include bond lengths, bond angles, ground state vibrational frequencies, ionization potentials, electron affinities, heats of formation, hydrogen bonding interaction energies, conformational energies, and reaction barrier heights. Finally, a list of the best-performing methods is given.

The final chapter gives a brief summary of the work presented in this dissertation. More complete discussions and conclusions are given within each chapter. The computational investigation of biological systems is necessary to completely understand the mechanisms by which macromolecules function within the body. The metalloprotein

studies presented here incorporate the use of a variety of computational tools to address questions that would be very difficult to answer using other types of experiments. The DFT survey is a prime example of the usefulness of powerful computer clusters, and provides a detailed comparison of commonly-used density functional methods. This research is made possible by the continuous development of computational methods and computer power and serves as a testament to the flexibility and far-reaching applicability of QM and MM calculations.

CHAPTER 2 QM AND MM METHODS OF THE STUDY OF METALLOPROTEINS

The behavior of metalloproteins can only be fully described using a combination of several computational techniques. For instance, MM methods do not take electrons into account, while high-level QM calculations are too computationally expensive to be tractable for large systems over 120 atoms or so. The electronic structures, atomic charges, and accurate geometries of metal binding sites should be investigated with quantum mechanical methods whenever possible. Data from these studies form a foundation for the molecular mechanics or semi-empirical methods applicable to large-scale systems.

This section discusses the various computational techniques utilized in the studies of HAH1 and AAP presented later in this thesis. A brief discourse on the background and theory of QM methods is given in chapter 5 as an introduction to the DFT study.

Metal Binding Site Studies in Gaussian 03

Creating models of metal binding sites in proteins for the purpose of QM calculations is a common practice in studies that aim to describe the metal binding environment.⁸⁻¹¹ For the study of the Cu(I) binding site of the Cu transport protein HAH1, a model active site was created from the X-ray crystal structure of the Cu(I) binding site in the HAH1 dimer¹ using thiol [HSCH₃] or thiolate [SCH₃]⁻ ligands instead of the native Cys residues to complex the central ion. Using DFT in Gaussian 03, the thermodynamics, bonding parameters, and electrostatics of the model systems were determined.

The QM calculations of the model clusters should be reliable so that the MM parameters derived from the QM data are also reliable. One approach to ensuring a stable QM structure is to use several initial geometries of the metal cluster to so to avoid getting trapped in a local minimum energy well.⁹ Another way in which the accuracy and reliability of calculations of Cu(I) is increased is the development of Cu(I)-specific basis sets. Olsson and Ryde^{9,12} modified a double-zeta basis function of Schäfer¹³ by the inclusion of diffuse p, d, and f orbitals (DZpdf). Table 2-1 lists the Ryde basis set used for the production QM calculations of the model Cu(I) clusters. The modification from the original Schäfer basis is the enhancement of the p, d, and f orbitals with exponents 0.174, 0.132, and 0.39, respectively. A modified 6-31G* basis function has been created by Pulay *et al.* that can also be used for Cu and other 1st row transition metals.^{14,15} The modified 6-31G* basis set was used to generate an initial structure for the use of the Ryde basis set, optimizing the Cu(I) cluster in a stepwise fashion.

The initial structures for the model clusters were created in WebLabViewer Pro¹⁶ from the Cu(I)-bound HAH1 dimer crystal structure 1FEE¹ in Cartesian coordinates and optimized in Gaussian using the default grid at B3LYP/m6-31G*, where m6-31G* is the modified basis set of Pulay *et al.*¹⁴ Since the HAH1 model clusters are all closed-shell singlet states, RB3LYP could have been specified in the Gaussian input file to request a restricted calculation. While this method is less computationally expensive, restricted and unrestricted calculations will produce the same results for singlet species. Further optimization was performed using the Ryde DZpdf basis set for Cu(I) and the split valence 6-311++G* basis for S, C, and H. Upon final optimization at this level, each of the clusters was subjected to single point calculations for frequencies and thermodynamic

properties. Bonding force constants between Cu(I) and S were also obtained. Gaussian outputs a user-readable table of bonds and associated constants based on internal coordinates.

Electrostatic potential calculations were also performed on the optimized structures. The Merz-Kollman-Singh charge method^{17,18} was used and the ionic radius for Cu(I) was set as 0.91 Å. The MKS method spreads the partial atomic charge (the calculated electron moment) of each atom onto the surface of a sphere around the atom defined by its ionic radius. Gaussian ESP's were converted to AMBER charges using the *espgen* and *respgen* programs in the AMBER suite. These three calculations account for all of the bonding and electronic properties of the model that are needed to create a MM force field for the native metal binding site for Cu(I)-bound HAH1.

The AAP system was also subjected to geometry optimizations and single point frequency calculations following the same procedures as above. Once again, the active site from a high-resolution crystal structure of the apo-form of AAP was used to create an initial structure model for the AAP active site. Residues were abridged and capped at their termini except in cases of adjacent residues, for which the peptide bonds were retained for the calculations. Since the creation of a new force field was not the goal for the AAP system, and the primary interests were the electronic structure of the active site and the parameters of Zn(II) ion coordination within the active site, the B3LYP/6-31G* method was sufficient for the calculations. Also, Zn had already been sufficiently described by the normal 6-31G* basis function, so no modified or larger basis sets were needed for the metal ions.¹⁵ Population analysis for atomic charges was not performed on

the AAP active site. The model for the AAP active site needed to be more robust than that of the HAH1 active site since the goal of the study was to elucidate the finer details

Table 2-1. The DZpdf basis function used for Cu(I) in the DFT calculations of the cluster models.

Orbital	Exponents		Orbital	Exponents	
S	6	1.00	P	3	1.00
	441087.2507	-2.18E-04		73.67182138	0.238814523
	66112.02119	-1.69E-03		30.44736969	0.449800159
	15047.01143	-8.81E-03		13.12271488	0.393376824
	4263.427308	-3.60E-02	P	1	1.00
	1396.38158	-0.117429705		5.521483997	1
	511.9605579	-0.288442674	P	1	1.00
S	2	1.00		2.145792213	1
	203.4542695	-0.426788989	P	1	1.00
	82.79233703	-0.330441285		0.767974887	1
S	1	1.00	P*	1	1.00
	20.85428563	1		0.174	1
S	1	1.00	D	3	1.00
	9.041067958	1		47.3137437	3.24E-02
S	1	1.00		13.15468845	0.168227065
	2.751813517	1		4.366288575	0.384944296
S	1	1.00	D	1	1.00
	1.043485652	1		1.412206594	1
S	1	1.00	D	1	1.00
	0.111722924	1		0.38840713	1
S	1	1.00	D*	1	1.00
	4.10E-02	1		0.132	1
P	3	1.00	F*	1	1.00
	2530.096567	1.91E-03		0.39	1
	600.0979295	1.58E-02			
	194.0820448	7.63E-02			

* Represents enhanced orbital exponents
Courtesy of Olsson and Ryde.⁹

of the electronic structure of the complete di-Zn(II) binding site to the 1st coordination shell. Therefore, the model AAP active site comprised about seventy atoms for 1st shell coordination sphere compared to 21-23 atoms for the HAH1 model cluster.

Molecular Mechanics of HAH1

The completion of the set of QM calculations on the HAH1 model site led to the next phase of the experiment, the description of the MM force field for the Cu(I) binding site. AMBER⁶ is a suite of programs that perform a variety of calculations and analyses based on user input and a collection of libraries. The research discussed here utilized AMBER versions 6 -9. Despite the many versions used, the force fields and libraries used in the course of this study remained the same. For the purposes of the calculations discussed here, the 1994 force field of Cornell *et al.*¹⁹ was used in conjunction with the 1994 libraries and the parameters constructed for Cu(I) and Cu(I)-bound ligands after the QM work described earlier. The 1994 force field is a large database of parameters detailing bond lengths, bond angles, bond torsions, nonbonding parameters, and electrostatics for dozens of specific atom types occurring in proteins, nucleic acids, and organic compounds. The parameters within the force field libraries are derived from QM calculations of organic molecules and amino acids. Each of these parameters contributes to the overall energy of the system being investigated.

The AMBER Force Field

The AMBER energy E_{AMBER} is defined as the enthalpy difference between the folded and unfolded states of a protein and can be calculated as the sum of the energy contributions of the parameters listed above. This expression is known as the AMBER force field (equation 2-1). The first three terms of 2-1 are bonding interactions. Bond stretching and bond angle bending contributions to the energy are calculated using force constants applied to each term. Equilibrium values and force constants are user-defined and the nature of the energy functions sets the equilibrium value at the bottom of a potential well. The third term involves the bond torsion contribution to the total energy

represented by a truncated Fourier series in which the measured torsion angle ϕ and the phase γ are input and V_n is the rotation barrier height for the torsion.^{20,21}

The last two terms represent the nonbonding interactions between atoms i and j . The fourth term is the 6-12 Lennard-Jones potential, accounting for the attraction and repulsion between two nonbonded atoms known as van der Waals interactions. The $\frac{1}{R^{12}}$ term accounts for repulsion, while the attraction is calculated in the $\frac{1}{R^6}$ term. In this term, A and B represent the lowest-energy distance between the two interacting atoms, and the term can be modified by the inclusion of an artificial energy well in the potential curve. The final term includes the Coulombic interaction between charged nonbonded particles. The magnitude of this interaction depends on the charges of the two interacting particles and the distance between them embedded in field with a dielectric constant, ϵ . In AMBER, a nonbonding cutoff can be defined to set a maximum distance at which these interactions can be felt between two particles.²⁰

$$\begin{aligned}
 E_{\text{AMBER}} = & \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \\
 & \sum_{\text{dihedrals}} \frac{V_n}{2} (1 + \cos[n\phi - \gamma]) + \\
 & \sum_{i < j}^{\text{atoms}} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \sum_{i < j}^{\text{atoms}} \frac{q_i q_j}{\epsilon R_{ij}}
 \end{aligned} \tag{2-1}$$

The difference between MM and QM methods is the inclusion of nuclear and electronic interactions in QM methods. Except for the electrostatic interactions in the last term of equation 2-1, no quantum effects are considered in the MM force field. Even

then, electrostatic potentials are user-defined or drawn from a library in AMBER, not explicitly calculated as with QM methods.

In essence, the molecular dynamics package in AMBER, called *sander*, relies solely on the AMBER potential to calculate all interactions between the particles in a system in a stepwise fashion. Several files must be prepared in order for a successful simulation or minimization to be performed in *sander*. The necessary files are created in the LeAP program within AMBER.

LeAP and *sander*

LeAP is initialized by loading a force field and parameter library suitable for the system being investigated. The Cornell 1994 force field is used throughout this work, which includes the parm94.dat force field and the all_amino94.in amino acid database. These two files form the basis for the atoms and residues that LeAP recognizes. Next, the initial structure file for the system under study is loaded into the interface. A check of the structure will output any faults within the system such as non-integral charge or unknown atom types. In the case of HAH1 MD and free energy simulations, Cu, Cu-bound S, and any dummy atoms had to be described in separate force field modification files (.frcmod) before LeAP could recognize them as atom types.

At this point, any new parameter files can be loaded into the interface and applied to the structure. Parameter files must be correctly formatted to be AMBER-compatible and may contain user-defined atom names, atom types, and many other atomic parameters. Furthermore, bond connections can be defined or broken and atomic partial charges can be specified for certain atoms if they are unknown or if they differ from the values in the AMBER library so that the total charge of the system is an integer. Incidentally, AMBER charges differ from any charges that may have been generated

outside of AMBER using a QM method like Merz-Kollman-Singh. Separate programs in AMBER, called *espgen* and *respgen*, are used to read the output from a QM charge calculation and convert the QM charges into AMBER charges. A successful check of the system within LeAP should reveal no further unknown atom types or electrostatic errors.

sander uses two files to describe the system. These are the coordinate file (.crd) and the topology file (.top). Once all of the parameters have been specified for the system in LeAP and there are no further errors, the .top and .crd files can be created. Now, LeAP will return any parametrical errors such as missing bond, angle, or torsion force constants or undefined atoms types within the system that need to be added or amended before the two new files are created.

sander reads one more file. This is the input file for the program. The first simulation performed on HAH1 was a MM minimization. Minimization routines vary, but the default in AMBER is the steepest descent (SD) method. This method calculates the gradient of the energy difference between two conformations of the system and initializes the next minimization step along a linear path so that the energy of the system is decreased most rapidly. After some number of steps, *sander* may convert to the conjugate gradient (CG) method of energy minimization to aid in convergence. CG differs from SD in that a more direct route to a minimum energy state is taken. CG is particularly well-suited for finding the minimum of a shallow potential surface or an elongated energy well. Many conditions of the minimization may be defined including minimization method and number of steps to be taken.

Although it is unlikely that the global minimum of a large system of many thousands of atoms will be found by the methods used by *sander*, a sufficiently

minimized structure will be generated after a large number minimization steps have been performed. Traits of a sufficiently minimized structure are low vdW, bond, angle, torsional, and electrostatic energies compared to the initial structure. The user can monitor the progress of a minimization run by tracking the decrease of energy components and by visually inspecting the system at any point during the process.

Once the gas-phase structure has been minimized, it can be equilibrated in the gas phase with MD or it can be solvated. The creation of an ordered “water box” is performed in LeAP. In these simulations, TIP3P water boxes were added to the gas-phase minimized and equilibrated proteins. The TIP3P box is a snapshot of an equilibrated water box,²⁰ and the use of actual water molecules around the protein instead of an external dielectric field is called explicit solvation. The protein fits within the box of user-defined dimensions so that there is plenty of water surrounding the protein. In this research an 8.0 Å or 10.0 Å water box was used to solvate the protein. The solvated HAH1 protein system comprises more than 19,000 atoms compared to about 2,050 atoms in the gas-phase system.

To protect waters on the edge of the simulated box from being exposed to the vacuum that exists outside of the box, periodic boundary conditions (PBC) are applied to the system. PBC overcomes the problem of the outer solvent molecules being directly exposed to vacuum by replicating the real box containing the solvent and solute and placing imaginary copies on all sides of the real box. That way, if a water molecule or part of the solute leaves the edge of the real box, a water molecule entering the other side of the box from an imaginary neighbor replaces it. Another feature implemented in MD simulations of solvated systems is Particle Mesh Ewald (PME), in which short-range and

long-range interactions are separated and calculated in different ways in order to decrease computational expense. Without PME, the generation of the list of nonbonding interactions like vdW and electrostatics for large systems would take a very long time.

The solute must equilibrate with the newly applied solvent. Constraints are imposed onto the protein within the box so that it does not change conformation as the water equilibrates around it. The temperature of the system is slowly raised to 300 K over thousands of equilibration steps, and the restraints on the protein within the solvent box are also gradually relaxed. Once the system reaches 300 K, the constraints are removed and the whole system is allowed to equilibrate. Once the solvated system is sufficiently equilibrated, it is subjected to long term MD simulation.

The prediction of molecular motion through time is derived from subjecting a system to Newtonian laws of motion in conjunction with a potential function, in this case the AMBER force field. The force felt by particles in the system is the derivative of the potential with respect to position.

$$\mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{r}_i} \quad (2-2)$$

Moving particles through time is done in iterations of calculating the incident force on each particle and then moving the particles in reaction to that force. With the initial positions of the particles known, the force is actuated and the particles are allowed to move. The positions of the particles after a small time step Δt can be calculated using the Verlet equation.

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \frac{\mathbf{F}(t)}{m} \Delta t^2 \quad (2-3)$$

Subsequent positions are formulated in the same manner until the user-defined number of steps has been performed. Each step results in a new conformation of the system that can be written to an output file. Typical MD step sizes range from 0.5 to 1.5 femtoseconds, and the total length of production-quality MD simulations is generally four nanoseconds (at least four million steps with a 1.0 fs time step). A collection of these individual snapshots can be analyzed to see how the system changes over time.

In the HAH1 system, the geometry of the active site was monitored to ensure that Cu(I) was always bound in the correct coordination and that the active site behaved well. This validated the new parameters that were imposed on the Cu(I) binding site. The root-mean squared deviation (rmsd) from the crystal structure was monitored to quantify how the simulated protein differed from the reference state during the simulation. Also root mean squared fluctuation (rmsf) of key atoms and residues were calculated to see which parts of the protein were highly mobile and which maintained more fixed positions. Radius of gyration (rad_{gyr}) is another way to quantify the motion of the system in a general manner. This can be interpreted as the rmsd of scattering elements from the particle's center of mass.

$$\text{rmsd} = \sqrt{\frac{\sum_{i=1}^{N_{\text{atoms}}} d_i^2}{N_{\text{atoms}}}} \quad (2-4)$$

$$d_i = \left(\mathbf{r}_{i,\text{sim}} - \mathbf{r}_{i,\text{extl}} \right) \quad (2-5)$$

Equation 2-4 calculates the rmsd between a reference state and an individual structure generated by the MD simulation. rmsd is probably the most basic function for quantifying the difference between two structures. Good rmsd values between a crystal structure reference state and a MD snapshot of the simulated protein are generally lower

than 3.5 Å. The movement of a subset of the system with respect to the average structure over the whole simulation is the rmsf, and the B-factor is related to rmsf by equation 2-7.

A high B-factor identifies a highly mobile subset.

$$\text{rmsf}_i = \sqrt{\langle r_i^2 \rangle - \langle \mathbf{r}_i \rangle^2} \quad (2-6)$$

$$\mathbf{B}_{AMBER} = \text{rmsf}_i \cdot \frac{8}{3} \pi^2 \quad (2-7)$$

Finally, rad_{gyr} is a measure of how much a protein spreads out from its center. In essence, it quantifies the level of unfolding of a protein. rad_{gyr} is size dependent, as larger proteins have larger rad_{gyr} values. It is also shape dependent and different kinds of structures (coil, sheet, random loop) all have different rad_{gyr} .

$$\text{rad}_{\text{gyr}} = \left\{ \frac{\sum_{j=1}^N \left[r_j - \left(\frac{\sum_{i=1}^N r_i}{N} \right) \right]^2}{N} \right\}^{\frac{1}{2}} \quad (2-8)$$

Each of these values can be obtained using another program in the AMBER suite called *ptraj*. This program reads in a reference state and the coordinate file of the MD simulation (which may contain thousands of conformations for a system that is thousands of atoms big) and monitors user-selected parameters for the entire MD simulation. *ptraj* is also used to generate structure files for individual poses.

Free Energy Simulations of HAH1

Once the integrity of the HAH1 parameters had been validated by analyzing the MD simulations, free energy calculations on the protein began. One of the goals of the

HAH1 study was to identify an energetically favorable transition state of Cu(I) transfer between HAH1 and its target MNK. Energetically, the enthalpy, entropy, and free energy of different Cu(I) coordination states could be determined. Of these three properties, free energy provides the most accurate account of the relative energies of different systems.

As with the MD simulations, an initial structure was constructed for the TI calculations. However, the reactions involved in the TI calculations add a level of complexity to this process. TI simulations perturb the system from one state to another. In this case, the initial structure was a three-coordinate Cu(I)-bound active site and the “product” of the perturbation was a four-coordinate Cu(I) active site in HAH1. These two conformations differ by one atom, which is the hydrogen on the unbound Cys of the initial structure. To compensate for the “disappearance” of that hydrogen, a dummy atom was created in the final state to ensure that the total number of atoms in each state was the same.

Once the number of atoms in the two states was equal, then the initial structure was loaded into LeAP. The initial structure used the same parameters as the MD simulation on the three-coordinate Cu(I) structure used. However, the four-coordinate product had different charges and bonds than the starting structure. These changes were outlined in LeAP and represented the final state of the TI simulation. In AMBER, the perturbed partial atomic charges and atom types were input for every atom that was changed during the simulation. Now, the parameters of both the initial and final states of the TI simulation have been described in LeAP. A check of the system should reveal no unknown atom types and that the total charges of the two states are integer values.

The final phase in LeAP for preparing free energy calculations is the generation of the coordinate files and topology files for the two states. The unperturbed topology file will be used to minimize and equilibrate the initial structure prior to the TI simulation. The perturbed topology file will be used in the actual TI calculation and contains all of the information for both states of the simulation. If there are any incomplete parameters pertaining to either state, LeAP will alert the user at this point. The perturbed topology file will only be generated when the perturbed system is fully described.

At this point, the initial structure is minimized and equilibrated in the same manner described above. Gas-phase and solvated protein structures are both equilibrated and prepared for TI simulations. Twelve *sander* input files are created; one for each window of the TI simulation. The **icfe=1** keyword turns on TI in *sander*. The details of the TI calculations are provided in chapter 3.

Several requirements must be met to ensure the accuracy of the TI calculations. A good potential function must be used to describe the system. In this study, the AMBER force field was used along with the parameters derived for the Cu(I) binding site. Secondly, there must be a way to rapidly update the system as it changes over the course of the free energy simulation. This includes evaluating forces and energies and updating positions through time. Finally, ΔG must be calculated. Equation 2-9 shows how the force is calculated as the second derivative of position with respect to time, and equations 2-10 and 2-11 reveal how ΔG can be determined in the canonical ensemble (constant N, V, and T). In equation 2-11, the Hamiltonian H is approximated by the AMBER force field. Although equation 2-9 allows rapid calculation of the force, it requires high accuracy. This explains why the time length of steps in MD and TI calculation should be

about 1 fs. In order to sample a sufficiently large conformational space, the number of steps in each window must be very large.²²

$$\mathbf{F} = -\frac{\partial V(\mathbf{r})}{\partial \mathbf{r}} = m\mathbf{a} = \frac{m\partial^2 \mathbf{r}}{\partial t^2} \quad (2-9)$$

$$G_{\text{NVT}} = -RT \ln Q_{\text{NVT}} \quad (2-10)$$

$$Q_{\text{NVT}} = C \int \int e^{-H(\mathbf{x}, \mathbf{p})/RT} d\mathbf{x}d\mathbf{p} \quad (2-11)$$

Unfortunately, these calculations take a very long time to converge even for simple systems such as the water dimer. Solving explicitly for G is, in essence, an expansion of H and higher orders of the expansion take even longer to converge. Therefore, other methods of calculating G must be derived using some approximations.

Thermodynamic Integration

Another functionality of the *sander* program is the ability to perform thermodynamic integration (TI) calculations. Like other free energy methods, TI perturbs an initial structure to a final structure over a series of windows. TI uses a scaling parameter λ , which varies between 0 and 1 as the system's character progresses toward the final structure. At $\lambda=0$, the system exists completely in the initial state, and at $\lambda=1$ the system exists completely in the final state. TI is based on the integration of the ln form of the G expression (equation 2-10) and where H is $V(\lambda, \mathbf{x})$.

$$\frac{dG}{d\lambda} = -\left[\frac{RT}{Q_{\text{NVT}}} \right] \cdot \left[\frac{dQ_{\text{NVT}}}{d\lambda} \right] \quad (2-12)$$

Now, substituting for Q_{NVT} :

$$\frac{dG}{d\lambda} = \frac{\int \left(\frac{dV(\lambda, \mathbf{x})}{d\lambda} \right) e^{-V(\lambda, \mathbf{x})/RT} d\mathbf{x}}{\int e^{-V(\lambda, \mathbf{x})/RT} d\mathbf{x}} = \left\langle \frac{dV(\lambda, \mathbf{x})}{d\lambda} \right\rangle_{\lambda} \quad (2-13)$$

And integration yields ΔG :

$$\Delta G = \int_0^1 \left\langle \frac{dV(\lambda, \mathbf{x})}{d\lambda} \right\rangle_{\lambda} d\lambda \quad (2-14)$$

Yet another obstacle presents itself here in that equation 2-14 is not analytically solvable and must be solved numerically with another approximation.

$$\Delta G \approx \sum_0^N \left(\left\langle \frac{\partial V(\lambda, \mathbf{x})}{\partial \lambda} \right\rangle_{\lambda(i)} + \left\langle \frac{\partial V(\lambda, \mathbf{x})}{\partial \lambda} \right\rangle_{\lambda(i+1)} \right) \times \left(\frac{\lambda(i+1) - \lambda(i)}{2} \right) \quad (2-15)$$

Or simply,

$$\Delta G \approx \sum_{i=1}^n w_i \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\lambda_i} \quad (2-16)$$

sander outputs $\frac{\partial V}{\partial \lambda}$ values. The user can define how often these values are

calculated and averaged. w_i is the weighting factor for each window. The TI calculations

used in this work contained twelve windows between $\lambda=0$ and $\lambda=1$. The $\frac{\partial V}{\partial \lambda}$ values from

the *sander* output were averaged for each window, multiplied by the weighting factor for

that window and then all of the weighted values were summed to generate the ΔG value

for the TI perturbation. Table 2-2 lists the weighting factors used the ΔG calculations. λ

values for the twelve windows are listed in Table 2-3. For the TI simulations of the

solvated protein, 1,000 $\frac{\partial V}{\partial \lambda}$ values were collected for each window (each of those $\frac{\partial V}{\partial \lambda}$

values represented 500-step averages within each window). The average of the 1,000 $\frac{\partial V}{\partial \lambda}$ values for each window were then averaged and weighted.

Placement of the windows along λ and weighting factors are based on Gaussian quadrature, which quantifies the integral of the space under the $\frac{\partial V}{\partial \lambda}$ curve with an accuracy similar to that of simpler methods like the midpoint method or the trapezoid method, but the Gaussian method requires only half the sample size of the simpler methods which saves on simulation time.

Table 2-2. Windows and weighting factors for a 12-window TI calculation in *sander*.

Window	Weight
1 & 12	0.02359
2 & 11	0.05347
3 & 10	0.08004
4 & 9	0.10158
5 & 8	0.11675
6 & 7	0.12457

Table 2-3. λ values for a 12-window TI calculation in *sander*.

Window	λ
1	0.0092
2	0.04794
3	0.11505
4	0.20634
5	0.31608
6	0.43738
7	0.56262
8	0.638392
9	0.79366
10	0.88495
11	0.95206
12	0.99078

Free Energy Perturbation

Free energy perturbation (FEP) differs from TI in that FEP is a more continuous representation of the change from $\lambda=0$ to $\lambda=1$. Instead of performing twelve separate calculations for windows of the TI simulation, FEP is a single calculation containing as many windows as the user wants. Each window within an FEP calculation contains a number of equilibration steps followed by a number of data-gathering steps. Another way to perform a FEP is to evaluate the trajectory of some initial state with the Hamiltonian of the perturbed state. This closely resembles the calculation performed on HAH1 where the structure generated by the initial Hamiltonian was evaluated using an alternate Hamiltonian that instituted some vdW exclusions. This is discussed in chapter 3.

Instead of perturbing atom types and charges as in the TI simulations, the perturbation in the FEP calculations for the HAH1 study is the exclusion of some vdW forces in the perturbed state. ΔG between two states A and B is defined in equation 2-17.

$$\Delta G = G_B - G_A = -RT \ln \left(\frac{Q_B}{Q_A} \right) \quad (2-17)$$

Q has the same form as in equation 2-11. In the general equation for FEP,

$$\Delta G = -RT \ln \left\langle e^{-\Delta H_{\lambda_i \rightarrow \lambda_{i+1}}/RT} \right\rangle_A \quad (2-18)$$

ΔH is the change in the Hamiltonian (approximated by the AMBER force field) from the initial state to the perturbed state. In the case of the HAH1 study, the difference between the two Hamiltonians should only be in the vdW term. Essentially, FEP spans the space of two physical endpoints 0 and 1 with an array of non-physical states in between, characterized by the value of λ .

$$\Delta G_{\lambda=0 \rightarrow \lambda=1} = \sum_{\lambda} \Delta G_{\lambda_i \rightarrow \lambda_{i+1}} \quad (2-19)$$

The *gibbs* program in the AMBER package was created to be able to perform FEP on large systems, with one advantage over TI in *sander* of being able to define dummy atoms in both states. Despite numerous attempts, I was never able to adequately simulate the HAH1 system with *gibbs*. After AMBER 7, *gibbs* was no longer developed or supported in AMBER.

Potential of Mean Force

Potential of mean force (PMF) calculations were performed on the HAH1 system to determine the energetics of the Cu(I)-S bonds within the active site. PMF between two bodies is a function of the distance between their centers of mass.²³ In AMBER, a harmonic potential is applied to a certain bond (or any other parameter), centered at an equilibrium value. Then, over a series of windows, the parameter is varied from an initial position to a final position. The interaction between the two particles involved in the PMF is monitored. W can be an expression of the free energy change with respect to coordinates and is calculated in equation 2-20.

$$W(q) = -k_B T \ln \pi(q) \quad (2-20)$$

Here, q is the coordinate and π is the probability of q obtaining a certain value. In an MD simulation, the values of q can be collected into bins and then analyzed as a histogram. The histograms for all of the windows are aligned with WHAM,^{24,25} a weighted histogram analysis program so that a PMF curve is constructed.²⁶ In order for a range of possible values of q to be adequately represented, a very large number of samples must be taken. In the case of the HAH1 system, the Cu(I)-S was varied from

about 2.8 Å to 2.0 Å over the course of twenty-three windows. Each window sampled 100,000 steps for a total of over two million samples taken for just 0.8 Å of coordinate space. A biasing potential is applied in the AMBER PMF to ensure that the simulation sufficiently samples the bond lengths of interest to the HAH1 study. The biasing potential is shown in equation 2-21.

$$U(q) = \frac{1}{2}k(q - q_0)^2 \quad (2-21)$$

q_0 is the equilibrium value that is of interest in the calculation. By assigning a high value to k , a large energy penalty has to be paid to any coordinate that is too far from q_0 . For the HAH1 PMF calculations a biasing potential of 2,000 kcal/mol was placed around an equilibrium distance of 2.19 Å. The PMF calculation was specified in the *sander* input files by specifying **nmropt=1** for the program to read in the biasing potential and calling a PMF parameter file which contained that information for each window individually. This represents a TI-method of performing the PMF calculation with *sander*. Unlike the TI simulation of changing the Cu(I) coordination, there is no chemical change between the initial and final states of the PMF simulation, only a conformational change in the bond length that is being adjusted in each window.

CHAPTER 3
COMPUTATIONAL STUDIES OF THE CU(I) METALLOCHAPERONE HAH1

Cu(I) Biochemistry and the HAH1 System

Cu Metallochaperones

Several transition metals have been implicated in important intracellular biological processes. These metals, including Fe, Co, Zn, Mo, and Cu, are involved in such central biological roles in part due to their ability to exist in multiple oxidation states *in vivo*.²⁷ For example, copper exists in both the cuprous and cupric states within the cell and functions as a catalyst in both. Cycling of copper ions between two oxidation states can catalyze the production of highly toxic hydroxyl radicals within the cellular environment that can result in damage to many intracellular macromolecules. This creates a potential problem within the cell: metal ions such as Cu(I) and Cu(II) are essential for normal cell behavior, yet the “free” existence of these ions in the cell is clearly toxic.

A group of metal-binding proteins labeled “metallochaperones” have been shown to bind transition metal ions in both prokaryotic and eukaryotic cells. Of particular interest are the chaperones involved in copper and zinc transport in human cells. In human cells, a number of these chaperones deliver Cu ions to other copper-binding proteins or organelles. Noteworthy is the fact that these chaperones are not anti-toxins. Instead, they act to sequester the ion and transport it through the intracellular environment. It has been found that the average concentration of intracellular “free” copper is on the order of 10^{-18} M, translating to less than one unbound copper ion per cell in the human body.²⁸ This low concentration has been attributed to the over-abundance of

moderate and strong copper-chelating sites including metallothioneins, vesicular sites, and specific copper-binding proteins.²⁹ Yet, with all the other potential copper-binding sites, specific copper chaperones are able to acquire the ions as they enter the cell and distribute them throughout the cell as needed. These include the human antioxidant protein (HAH1), the Menkes and Wilson's ATPases, the human copper chaperone for superoxide dismutase (hCCS), and the human copper, zinc superoxide dismutase (SOD1).

Several different copper-transport routes within the cell are responsible for copper homeostasis to ensure that the total copper concentration (normally in the micromolar range²⁹) in the cell does not get too high or too low.^{27,29-36} This involves regulation of the amount of copper entering and exiting the cell. Copper must then be delivered from the trans-membrane proteins to the metal-binding sites of the correct proteins in the cytoplasm. Finally, the copper ions must be transported through the cell to the proteins and organelles where they are needed. Proteins that perform each of these functions have been discovered and studied. Defects in the metabolism of intracellular metal ions result in a vast array of health problems. For example, problems within the copper-transport structure of the cell results in Menkes Syndrome, Wilson's Disease, familial amyotrophic lateral sclerosis (fALS) disorders, and Alzheimer's disease.³⁰

Individual pathways and binding interactions will be discussed later, but a brief overview of metal binding and inter-protein transport is given here. According to Rosenzweig, copper binding proteins deliver the copper ions to their targets via "direct protein-protein interaction".³⁷ Moreover, copper transport between a chaperone and its target is thought to progress through the formation of a series of multi-coordinate

complexes until the ion has been completely released from the donor and bound by the receptor.^{1,37} With this in mind, another concept suggests that the chaperone donates its metal ion in an enzymatic fashion, lowering the energy barrier for inter-protein ion transport.^{29,31}

Recently, crystal structures have been solved for many Cu(I) transport proteins³⁸⁻⁴¹ and some transport mechanisms have been suggested.^{1,31,33,35,37} However, the exact mechanisms by which copper delivery is accomplished is still under intense study. Several key issues should be addressed when considering this problem. For example, the specificity of the donor-target interaction must be understood. Currently, highly conserved secondary structures between the donor and target at the metal binding site are believed to explain the problem of recognition. Possible protein rearrangement during metal transfer must also be investigated. Disulfide bonds near the binding sites may contribute to rearrangement of the fold during metal transfer. Protein docking should also be addressed. It has been suggested that docking involves electrostatic interactions²⁷, and that heterodimers or even higher order oligomers⁴² may be formed during copper transfer between proteins.

The recent structure determinations of many of the proteins involved in copper transport in both apo- and holo- forms have opened this field to computational study. Ideally, using computational tools, an investigator could model the binding sites of several proteins, ultimately using molecular dynamics (MD) simulations to determine the binding and transfer mechanisms for the processes described above. Some *ab initio* modeling of Cu binding to sulfur and Cys groups has been performed.^{8-10,38} Currently,

ZINDO⁴³, PM3(tm), *ab initio*¹⁴, and SIBFA12³⁸ are some of a limited number of methods for which Cu parameters have been established.

Cu(I) is generally unstable in aqueous solutions. However, it may be stabilized by sulfur-containing ligands or immobilized by a protein.⁴⁴ Generally, the copper-binding active site within a protein is characterized by the use of two or more Cys or His ligands for direct binding less than 2.5 Å from the ion, with Met residues or charged amino acids as part of the supporting structure at 3.5 Å – 8.0 Å away from the metal. The human copper, zinc, superoxide dismutase (SOD1) employs four His at its Cu(I) active site. Three His appear to be tightly bound (2.0-2.12 Å) while the other is bound to a lesser extent (3.12 Å). Each His is bound to Cu(I) by the δ_1 or ϵ_2 N of the imidazole ring. Cys residues are present at the binding sites of the human copper chaperone for superoxide dismutase (hCCS), HAH1, yeast antioxidant protein (Atx1), the yeast homolog to Menkes disease protein (Ccc2), and the Menkes disease protein (MNK4). These chaperones bind the ion with multiple Cys residues as part of a common MT/CXXC binding motif, which forms part of a turn between an α -helix and β -sheet. Some Cu(II) binding proteins such as human nitrite reductase and the plant electron-transfer protein plastocyanin employ both Cys and His at their active sites. This study focuses on optimizing structures for the Cu(I) binding sites in HAH1 as well as other multi-coordinate Cu(I) structures, and to observe the differences between two-, three-, and four-coordinate Cu(I) complexes.

Cu pathways within the Cell

Once through the membrane, copper must be delivered to other sites in the cell where it is needed. Metallochaperones perform this task. Due to the specificity of the

binding and transfer mechanisms between chaperone and target, there is a different chaperone for delivering copper to each target. Chaperones are grouped according to the primary structure of their binding site(s). Atx1-type proteins exhibit the MT/HCXXC³⁷ or MXCXXC²⁹ binding motif at the N-terminus. In yeast, Atx1 is responsible for delivering Cu(I) to Ccc2, the yeast CPx-type copper ATPase, for eventual incorporation into Fet3,²⁷ an important protein in iron metabolism. HAH1 delivers Cu(I) to domain 4 of the Wilson's disease P-type ATPase, MNK4, in human cells. Another pathway involves the chaperone for superoxide dismutase in yeast, yCCS or Lys7, and in humans, hCCS. The function of these proteins is to deliver Cu(I) to the copper, zinc superoxide dismutase, SOD1. This pathway is more complicated than the Atx1 pathway since the proteins involved contain multiple domains and the Cu(I) binding sites are more complex and are not located near the surface of the protein.

Atx1-like copper chaperones bind the Cu(I) ion in a loop between an α -helix and a β -sheet near the exterior of the protein. The binding residues are all Cys, although there may be some contributing electrostatic interactions to the binding site from nearby Met, His, and Thr residues. HAH1 exists as homodimer in the crystal structure, binding one Cu(I) per dimer. Each monomer donates up to two Cys residues for ion binding. Atx1 exists as a monomer, coordinating the Cu(I) in a two- or three-coordinate system in the binding loop described above. The targets for Atx1 and HAH1, Ccc2 and MNK4, respectively, share secondary structure homology with their chaperones. Thereby, the chaperone and target are able to dock and ion transfer can occur. A representation of the Cu(I)-bound HAH1 dimer is shown in Figure 3-3.

The CCS proteins are more complex, and their structures have not yet been fully resolved crystallographically. yCCS is a 27 kDa two-domain monomer that exists as a 54 kDa homodimer protein⁴². Each domain possesses its own unique copper binding site. hCCS is a three domain 274 residue monomer that exists *in vivo* as a 548 residue dimer. Once again, each domain shares structure homology with another chaperone in the cell. Both yCCS and hCCS domain 1 have similar Cu(I) binding sites as Atx1 and HAH1, namely the MT/HCXXC or MX/CXXC binding motif. Domain 2 in both yCCS and hCCS has similar folding to SOD1, although each lacks several key structural features of SOD1, differentiating them from each other. It should be noted here that SOD1 exists as a 32 kDa dimer *in vivo* and utilizes four His residues as its Cu⁺ binding site⁸. Finally, domain 3 in hCCS is a small 39-residue feature containing a CXC binding motif. It is believed that this domain is involved in the physical transfer of Cu(I) from hCCS to the Cu binding site in SOD1.

Of these pathways, copper transfer between CCS and SOD1 is perhaps the process that has been investigated the most. Of interest is the intra-protein transfer of Cu(I) between domains in CCS and the transfer of Cu(I) from the CXC binding site of hCCS domain 3 to the quad-His binding site in SOD1. It has been suggested that the transfer involves the formation of a heterodimer or even higher order oligomers between monomers of CCS and SOD1.⁴²

Copper Homeostasis

One of the fundamental problems in biological coordination chemistry today is the insufficient understanding of how metal ion concentrations are mediated by intracellular processes.³² On one hand, enough metal ions must be present with the cell to facilitate essential biochemical functions. However, transition metals, and copper in particular, are

prone to cause problems due to their catalytic nature and the presence of so many favorable metal binding sites. As mentioned, Cu can readily cycle between two oxidation states, which can catalyze the production of toxic radicals within the cell. Moreover, several amino acids can easily bind Cu ions, creating an abundance of copper chelation sites within the cell.

Copper chaperones are proteins that bind Cu ions in both the Cu(I) and Cu(II) oxidation states and perform a twofold purpose in maintaining Cu homeostasis. First, the metal-binding site on the metallochaperones must be able to bind Cu ions more readily than the other favorable yet non-essential Cu binding sites throughout the cell. Secondly, the chaperones must act to sequester Cu ions from the intracellular environment, or at least ensure that the Cu ions are always bound within another essential Cu-binding protein.

Although there are numerous Cu pathways within the cell,^{27,36} the copper chaperones involved in these pathways can be divided into two groups: trafficking proteins and metalloregulatory proteins.³² Trafficking proteins are confined to cell membranes and the cytoplasm and include trans-membrane metal transport proteins and water-soluble Cu transport proteins that exist in the cytoplasm, delivering metal ions to specific intracellular target proteins. Metalloregulatory proteins bind ions in a more permanent fashion, using the ions to regulate gene expression and cell function.^{32,45,46} Initially, Cu pathways were identified but the details of metal binding were unknown. A large number of crystallographic and spectroscopic studies in the last decade have clarified the details of Cu binding in both Cu-transport and Cu-regulatory proteins. One facet of Cu homeostasis that remains largely unresolved is Cu transfer between proteins.

Isolating membrane-bound proteins for crystallization is a daunting task for crystallographers. Further complication arises in producing crystals of metal-bound proteins. While no structures currently exist for the Cu-binding trans-membrane protein hCtr1, the structures of soluble holo-proteins HAH1 and Atx1 have been determined by NMR² and X-ray crystallography.^{1,47} The structure of the fourth domain of the Menkes disease protein MNK4 (the target for Cu(I) transport from HAH1) has also been determined by XRC⁴¹, and the interaction between HAH1 and MNK4 has been investigated.⁴⁸ The copper-transport complex of yCCS (and its human homologue hCCS) and SOD1 has also been well-characterized by X-ray crystallography.^{39,42,49-51}

Quantum Chemical Characterization of the Cu(I) Binding Site from HAH1

The work reviewed in this chapter focuses on Cu(I) transfer from HAH1 to the Menkes disease protein. This pathway was selected due in part to the availability of high-resolution structures of the Cu(I) donor HAH1 and the target domain MNK4. The Cu-binding domains of both the donor and target employ two Cys residues in an MT/CXXC motif to hold Cu in a multi-coordinate state. HAH1 exists as a dimer in solution, with each monomer containing one binding domain. The coordination state of the Cu ion in the dimer is yet unknown. EXAFS studies of holo-Atx1 suggest a three-coordinate state, with two Cys residues binding tightly to Cu at 2.25 Å and a third less strongly bound Cys at 2.40 Å which may be from an adjacent Atx1 molecule.⁵² The 1.80 Å resolution X-ray crystal structure of the Cu(I)-bound HAH1 dimer reveals four Cys in close proximity to Cu(I). This structure suggests a roughly tetrahedral geometry for Cu(I) with three strongly bound Cys at 2.30 Å and the fourth Cys at 2.40 Å. The coordination environment for MNK4 is believed to be similar to that of the HAH1 monomer or Atx1.

Cu(I) transport between HAH1 and MNK4 is thought to progress via a multi-coordinate mechanism.^{1,47,52} Figure 3-1 characterizes the proposed mechanism for Cu(I) transfer as a series of Cu(I)-S bonds forming with the target domain as Cu(I)-S bonds break within the donor. The energetics of this process have yet to be determined, and it is not known if a potential four-coordinate intermediate exists as part of the transfer mechanism. The order of Cu(I)-binding and release is also unknown. In the HAH1 monomer, Cu(I) is bound by Cys12 and Cys15. As the target domain comes into close proximity, one of these two residues releases Cu(I) first. In a similar manner, the target domain must also sequentially form bonds, but it is not known whether Cys14 or Cys17 of MNK4 is the first to complex the incoming Cu(I) ion.

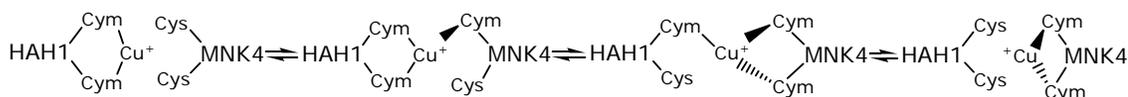


Figure 3-1. The proposed mechanism for Cu(I) transfer from HAH1 to the fourth domain of the Wilson's disease protein. Cym indicates a negatively charged Cys residue.

In order to address the details of Cu(I) transport, several models of the HAH1 Cu(I)-binding site were constructed in WebLabViewer Pro¹⁶ substituting Cys residues with methylthiolate [SCH₃]⁻ ligands. Two-, three-, and four-coordinate models were constructed and geometrically optimized using Gaussian 98.⁵³ Figure 3-2 depicts the optimized structures of the models and Table 3-1 lists some geometrical parameters of the structures.

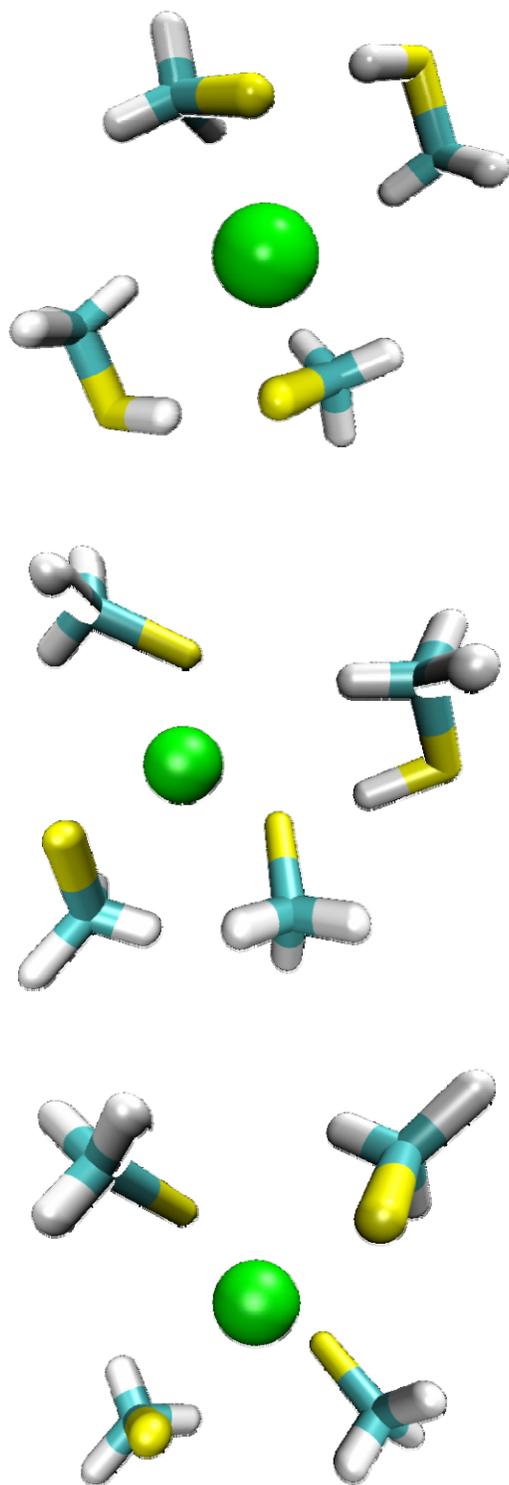


Figure 3-2. Gas-phase optimized structures of the multi-coordinate models of the HAH1 Cu(I) binding site. Cu(I) is in green, S is yellow, and C is gray. The top figure shows the two-coordinate model, followed by the three-coordinate model and the four-coordinate model at the bottom.

The structures were optimized using the B3LYP density functional combined with the Ryde double-zeta basis (DZpdf) set for Cu(I)⁹ and the 6-311++G** split-valence basis set for all other atoms using the six Cartesian-type d-orbitals.

Table 3-1. Geometry parameters of the DFT-optimized multi-coordinate models shown in Figure 3-2.

Model	Cu-S (Å)	Cu-S	Cu-S	Cu-S	S-Cu-S (deg)	C-S-Cu
2-coord	2.23	2.23	5.25	6.00	180.0	101.8
3-coord	2.31	2.35	2.41	4.98	114.6	105.2
4-coord	2.19	2.19	2.19	2.24	109.5	108.9

Upon optimization, the structures were used in single point calculations to determine atomic charges, thermodynamic properties, and bond force constants. Quantum charges were determined using the Merz-Kollman-Singh^{17,18} method. For these calculations, the Cu(I) ionic radius was set at 0.91 Å. Thermodynamic properties of the model clusters were obtained at the same level of theory as the geometry optimizations were performed. Finally, bond force constants were determined using the z-matrix form of the optimized geometry as the input.

The geometrical parameters, bond force constants, and MKS charge data were all used in conjunction to create a molecular mechanics (MM) force field for the Cu(I) binding site of HAH1.

Creation of a MM Force Field for the Cu(I) Binding Site of HAH1

The use of the HAH1 dimer as a model for the HAH1-MNK4 heterodimer

The X-ray structure of HAH1 shows that the protein crystallizes as a homodimer,¹ but NMR structures show that HAH1 exists as a monomer in solution.² The target protein, domain 4 of MNK, interacts with the HAH1 monomer in the cell, forming a heterodimer during Cu(I) transfer. Unfortunately, no structures exist (NMR or X-ray) for

the HAH1-MNK4 dimer or between HAH1 and any other MNK domain. Such a structure would be useful in MM and QM studies on Cu(I) transfer. Instead, an acceptable homolog to the HAH1-MNK4 heterodimer must be employed in such studies. Arnesano and coworkers performed a docking study of HAH1 to MNK4 to investigate the interactions between the metal binding domains and the protein-protein interface of the donor and target proteins.⁵⁴ Docking of the yeast antioxidant protein ATX1 to its target Ccc2 was performed in order to map the protein-protein interactions that facilitate Cu(I) transfer between the two proteins. Superimposition of the docked yeast heterodimer onto the crystal structure of the HAH1 dimer revealed that the two structures could be considered “remarkably similar”.⁵⁴ Larin and coworkers performed a manual docking of HAH1 to MNK4.⁴⁸ This study was performed previous to the elucidation of the NMR or X-ray structures of HAH1 and the HAH1 dimer, although the X-ray structure of MNK4 has already been resolved.⁴¹ In the Larin study, the homology between MNK4 and HAH1 was known, so one MNK4 domain was computationally adapted to model HAH1 in a computationally docked HAH1-MNK4 heterodimer. The information provided by these two studies suggests that the use of the HAH1 homodimer as a computational mimic for the HAH1-MNK4 heterodimer is a valid approximation.

Creation of parameter files for Cu(I)-bound HAH1

Cu(I) and S bound to Cu(I) are not defined atom types in the current version of the AMBER suite.⁶ AMBER 6 and AMBER 7 were used for the bulk of this work. In order to perform molecular dynamics simulations on Cu(I)-bound HAH1 and MNK4, these atom types must be defined in a format that AMBER can understand. This involves creating a force field parameter file that includes all the pertinent information used in the AMBER force field equation. For the purpose of this study, force field parameters

included molecular mass, two bond lengths types: Cu-S and S-C and their force constants, numerous bond angle types: S-Cu-S, C-S-Cu, C-C-S, and H-S-Cu and their force constants, a host of torsion angles and torsion constants, and van der Waals radii for Cu(I) and S. Initially, Cu-S bond lengths and bond force constants were implemented directly from the QM calculations on the model systems. C-S bond parameters in the metal-binding site were taken from the 1994 force field as were bond angles, angle force constants, and torsion parameters. Cu-S bonding parameters were varied from the QM-derived values after initial MD simulations those parameters revealed that the bonds were not strong enough to hold the desired binding site geometry. Once the geometry of the binding site was sufficiently described, the atomic charges were added. Cu(I) changes the normal charges of the adjacent atoms from their normal values in AMBER. Using the MKS charges output by the Gaussian calculation mentioned earlier, the *antechamber* package in AMBER was used to generate the electrostatic potential (RESP) charges for use in the AMBER input package LeAP. Charges on bound Cys residues were modified from typical AMBER charges for the HAH1 system to compensate for Cu(I) binding in order to ensure integral charge of the system.

Cu(I)-binding residues were specified as the CYM residue type in AMBER. This form denotes a ten-atom negatively charged cysteine ligand. Conversely, unbound cysteines or other cysteines elsewhere in the protein are defined using the CYS residue type. The CYM side chain is defined $-C_{\beta}H_2S^{-}$ while the CYS side chain is $-C_{\beta}H_2SH$. Atom types for Cu(I) and copper-bound S atoms also had to be defined. Copper was defined as atom PP. The four S atoms were identified as SA, SB, SC, and SD.

Tables 3-2 through 3-4 list the parameters defined in the AMBER force field file for the HAH1 Cu(I) binding site. Table 3-2 gives atomic mass and van der Waals parameters for S and Cu(I). Table 3-3 shows bond lengths, bond angles, and force constants for each, and Table 3-4 lists the RESP charges used for CYM ligands and Cu(I). All Cu(I)-bound S atoms have been kept equivalent, as have bound and unbound residues in terms of atom types and charges. Cu-S bond force constant values in Table 3-3 were increased by a factor of five over the force constants generated by the Gaussian 03 calculations performed on the model systems. This adjustment was implemented after the initial force constant parameters were found to be not be strong enough to keep the Cu(I) ion bound within the active site. CT-S-Cu angle force constant parameters were based on the CT-S-H parameters for normal Cys from the parm94 force field. The parameters described here closely match those determined in a similar study for Cu(II) bound to Met and His by Comba and Remenyi.⁵⁵ Figure 3-3 shows the active site region of the high resolution crystal structure of Cu(I)-bound HAH1. This site was reproduced in AMBER for MM simulations of HAH1. Cu(I) and the ligating Cys residues from HAH1 are shown. Cu(I) is in green, and the Cys residues are shown in stick form pointing toward Cu(I) creating a nearly tetrahedral binding environment. The two top Cys residues are Cys 12A and Cys 12B that are more solvent-exposed. Cys 15A and Cys 15B reside close to the monomer interface and have less contact with solvent. The total charges of the different entire-protein models are: unbound=0; two-coordinate=-1; three-coordinate=-2; and four-coordinate=-3. The holo-HAH1 dimer comprises between 2059 and 2061 atoms depending on the number of coordinating CYM ligands, and includes C- and N- terminal caps on each of the monomers present in the structure.

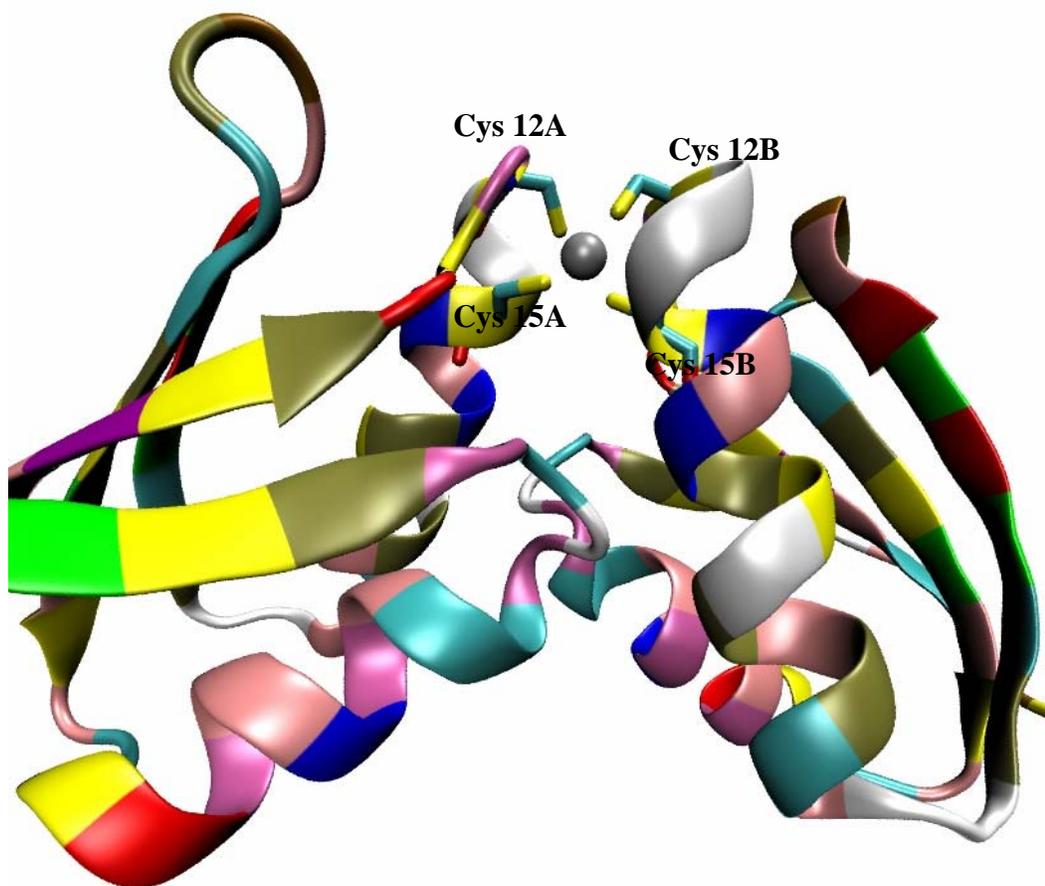


Figure 3-3. The 1.80 Å crystal structure for Cu(I)-bound HAH1. PDB ID 1FEE.¹

Once the force fields are fully described and the RESP charges are in place, the protein structure is ready to be minimized. The minimization and equilibration process takes several steps. First, the protein is described in LeAP and checked to ensure that all the bonds, angles, torsions and nonbonding parameters are fully described. Counterions can be added (Na^+ ions) to adjust the overall charge of the system to zero. An initial gas-phase minimization is performed on the protein. Then, the temperature of the system is gradually raised to 300 K over a series of MD runs. Once the temperature of the system reaches 300 K, it is subjected to a long MD run allowing the system to reach equilibrium.

Table 3-2. Atom type, atomic mass, van der Waals radii, and van der Waals well-depths for Cu(I) and Cu(I)-bound S in HAH1.

2-coordinate structure				
Atom	Atom type	Mass (au)	van der Waals radius (Å)	Well-depth (ε, kcal)
Cu	PP	63.55	2.50	0.20
S (Cys 12A)	SA	32.06	2.00	0.25
S (Cys 15A)	SB	32.06	2.00	0.25
3-coordinate structure				
Cu	PP	63.55	2.50	0.20
S (Cys 12A)	SA	32.06	2.00	0.25
S (Cys 15A)	SB	32.06	2.00	0.25
S (Cys 12B)	SC	32.06	2.00	0.25
4-coordinate structure				
Cu	PP	63.55	2.50	0.20
S (Cys 12A)	SA	32.06	2.00	0.25
S (Cys 15A)	SB	32.06	2.00	0.25
S (Cys 12B)	SC	32.06	2.00	0.25
S (Cys 15B)	SD	32.06	2.00	0.25

Table 3-3. Bond lengths, bond angles, and associated force constants for the HAH1 Cu(I) binding site.

Bond	k_{bond} (kcal/mol·Å ²)	r_0 (Å)
Cu-S	60.00	2.19
Angle	k_t (kcal/ mol·rad ²)	θ_0 (deg)
S-Cu-S	50.00	109.5
C-S-Cu	93.98	95.91

Table 3-4. CYM and Cu(I) RESP charges used for the HAH1 Cu(I) binding site.

Atom	CYS charge	Atom type	RESP charge in CYM		
			2-coordinate	3-coordinate	4-coordinate
N	-0.4157	N	-0.4408	-0.4157	-0.3630
H	0.2719	H	0.2468	0.2719	0.2520
C _α	0.0213	CT	-0.1000	-0.0351	0.0350
H _α	0.1124	H1	0.0257	0.0508	0.0480
C _β	-0.1231	CT	-0.0646	0.0168	-0.5720
H _β	0.1112	H1	0.0445	0.0053	0.2440
S _γ	-0.3119	SA, SB, SC, or SD	-0.8290	-0.8682	-1.0920
C	0.5973	C	0.6016	0.5973	0.6160
O	-0.5679	O	-0.5636	-0.5679	-0.5040
H _γ	0.1933	HS	n/a	n/a	n/a
Cu	n/a	PP	0.5922	0.6484	1.3670

The equilibrated system can be used directly for solvent-phase MD simulations. In order to perform MD simulation in solvent, a few more steps must be taken. The gas-phase minimized and equilibrated structure is again loaded into the LeAP program, where an explicit solvent box is added. In this case, an 8.0 Å TIP3P water box was imposed around the protein, increasing the total number of atoms in the system to over 19,000.

The solvated system was subjected to a series of relaxation runs similar to the gas-phase system. Incrementally smaller constraints were placed on the protein part of the solvated system as the system's temperature was increased to 300 K. Eventually, an equilibrated, solvated protein system was obtained. At this point, long timescale MD was performed. Each of the five multi-coordinate solvated protein model systems were subjected to the minimization and equilibration scheme just described. The solvated systems were ultimately simulated over a timescale of at least 3.6 ns.

Table 3-5 compares the AMBER-equilibrated HAH1 Cu(I) binding sites of the four-coordinate protein model to its model QM cluster counterpart minimized in Gaussian and to the X-ray crystal structure of Cu(I)-bound HAH1, and Table 3-6 lists other key data taken from the long timescale MD simulations. Although the geometries are not exactly reproduced by the MM force field parameters created for the HAH1 active site, the shape of the active site and the local protein environment are good. Figure 3-4 shows the root-mean-square deviation from the crystal structure for the entire protein sequence for each of the three solvated model systems, and Figure 3-5 displays the rmsd values for the active site loop regions of each protein model.

The figures below show that the active site regions of all five models reached equilibrium rapidly after about 400 ps. And while the entire proteins are in good agreement with the crystal structure throughout, the complete protein structures did take longer to reach equilibrium. For the whole-protein models, rmsd values of between 2.0 Å and 2.5 Å were achieved by 2,500 ps and were maintained beyond that point in the simulations. The rad_{gyr} values are within the expected range, and reveal that the most highly mobile residues are ones near the termini of the monomers.

Table 3-5. Comparison of the HAH1 active site between the four-coordinate model, the solvated, equilibrated protein, and the X-ray crystal structure of the Cu(I)-bound protein (1FEE).

Parameter	QM Model	Protein	X-ray
Cu-S (Cys 12A)	2.19 Å	2.29 Å	2.30 Å
Cu-S (Cys 15A)	2.19 Å	2.14 Å	2.39 Å
Cu-S (Cys 12B)	2.19 Å	2.33 Å	2.30 Å
Cu-S (Cys 15B)	2.24 Å	2.39 Å	2.32 Å
Cys12A-Cu-Cys15A	109.0 deg	117.5 deg	115.7 deg
Cys12A-Cu-Cys12B	109.5 deg	112.5 deg	109.4 deg

Table 3-6. Summary of rms deviation, rms flexibility, and radius of gyration for the five solvated HAH1 protein models and key active site residues.

	2-coord	2-coord (cis)	12B bound 3-coord (B)	15B bound 3-coord (A)	4-coord
	RMSD (Å)				
Total	2.69	2.67	2.27	2.10	2.04
Backbone	1.89	1.88	1.30	1.34	1.17
Binding loop	0.89	1.29	0.87	0.94	1.31
Bind. lp.	0.34	0.71	0.37	0.38	0.54
	Rad_{gyr} (Å)				
Protein avg.	29.38	27.26	29.35	29.30	29.43
	RMSF (Å)				
Total	4.98	4.85	3.60	4.54	5.46
Backbone	4.80	4.70	3.46	4.37	5.23
Cu	5.11	2.66	2.19	4.19	3.94
Cys 12A	5.28	3.83	2.40	4.95	5.08
Cys 15A	4.10	2.66	1.78	4.38	3.45
Cys 12B	6.14	3.62	3.44	3.64	4.68
Cys 15B	4.32	2.15	1.88	2.60	3.21

The rmsf values listed in Table 3-6 reveal details about the flexibility of certain residues as well as the complete protein and the protein backbone for the three models. The small difference between rmsf values for the complete protein compared to the backbone suggests that the flexibility of the protein is not limited to the side chains and that the backbone also move freely. From the rmsf data for the Cu(I)-binding residues Cys 12A, Cys 15A, Cys 12B, and Cys 15B, it appears that Cys 12A and Cys 12B have comparable magnitudes in each model. The values for Cys 15A and Cys 15B are also similar for each model. The similarity is derived from the location of these residues on the binding loop. Cys 12A and Cys 12B are more solvent-exposed and move more freely due to solvent interactions and being further away from the monomer interface. On the other hand, Cys 15A and Cys 15B show less flexibility as they are close to the interface region and not generally solvent-exposed. The flexibility of the Cys 12 residues may play a role in Cu(I) transfer between binding domains. rmsf data for Cu(I) show that Cu(I) is least mobile when bound by only three residues. In the four-coordinate model, Cu(I) is more flexible. This may be a result of the Cu(I) moving around within the binding site as different binding ligands move in and out of proximity to the ion. Perhaps, Cu(I) maintains a three-coordinate state even in the four-coordinate model, but complexes different residues over time.

The results from the MD simulations of the three Cu(I)-bound HAH1 dimer models show that the QM-derived parameters used to construct the MM force field adequately described the system. The rmsd data show that the computationally generated structures maintain the same fold and Cu(I) binding affinity as the protein *in vivo*. After the MD simulations were completed, the question of deciphering the order of Cu(I) binding and

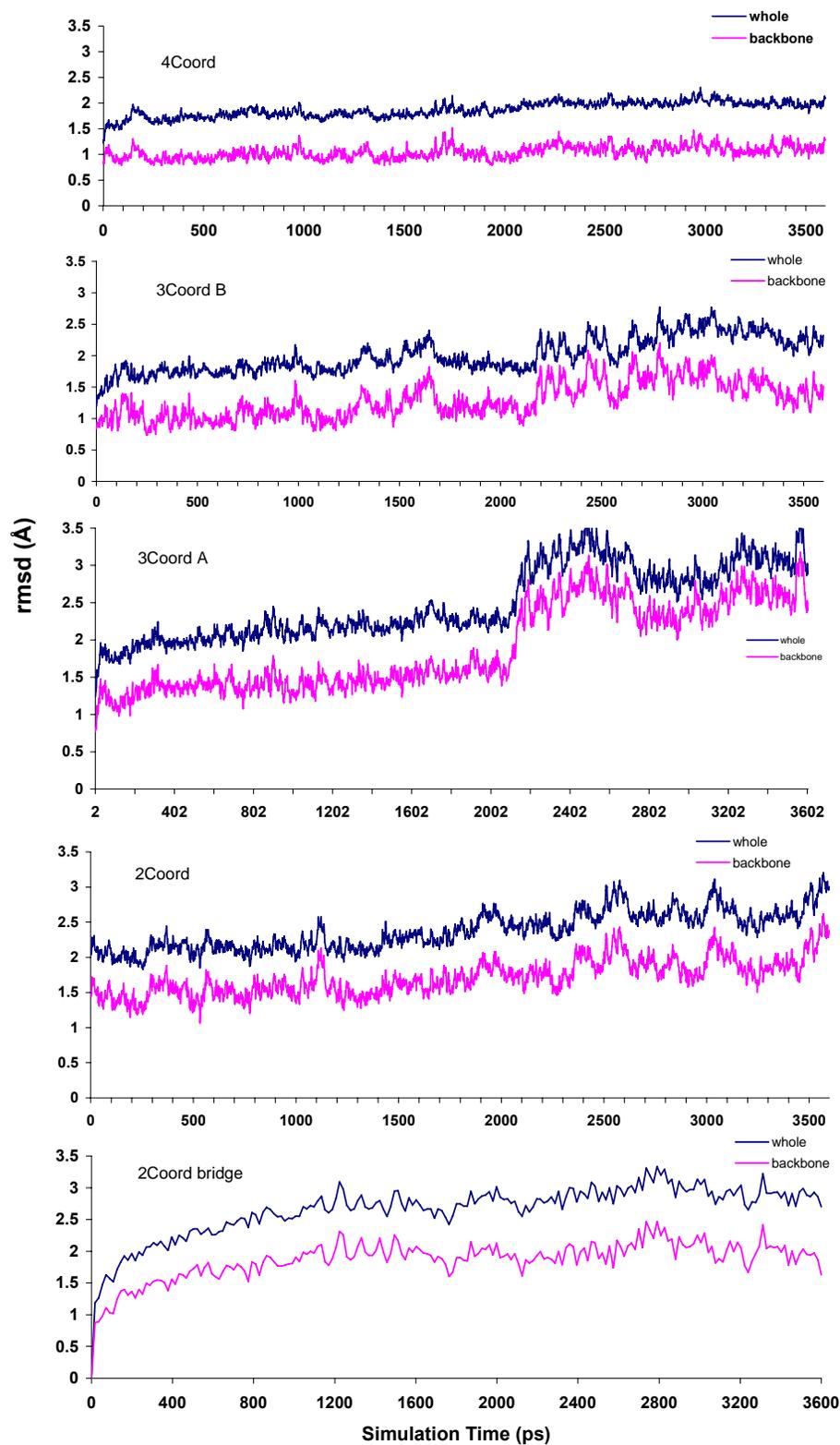


Figure 3-4. Root-mean-squared deviations between the five solvated HAH1 models and the Cu(I)-bound HAH1 crystal structure as a function of time.

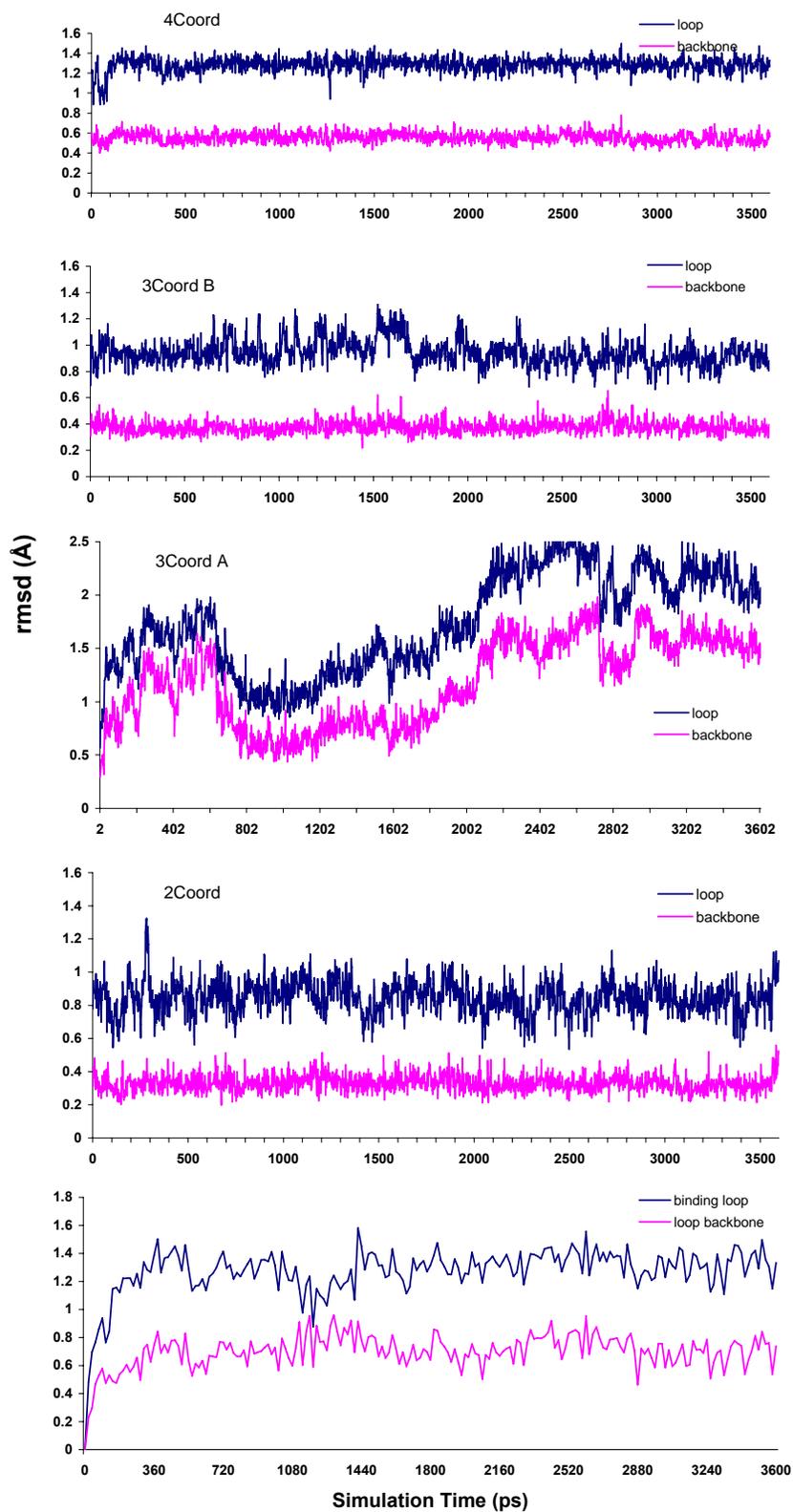


Figure 3-5. rmsd between the active site loop regions of the five solvated HAH1 models and the Cu(I)-bound HAH1 crystal structure as a function of time.

release during Cu(I) transfer remained to be answered. As the next section describes, the thermodynamic details of Cu(I) transfer were elucidated through a series of free-energy calculations using several different methods in AMBER.

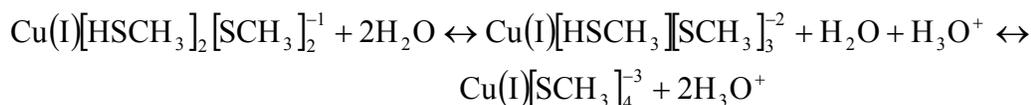
Free Energy Calculations of the Cu(I)-Bound HAH1 Dimer

Since the HAH1 Cu(I) binding motif has been described using QM and MM methods, the data garnered from those studies can be used to get an idea of the thermodynamics of Cu(I) transport so that more can be understood about the mechanism through which Cu(I) is transferred between the active sites of two metalloproteins. This study focused on proposing the order of Cu(I) binding and release based on an energetically favorable Cu(I) transfer pathway. As mentioned, the currently proposed mechanism suggests that Cu(I) is handed off from the donor to the target via a series of multi-coordinate Cu(I) intermediates during which the coordination number of the Cu ion is no lower than two.^{1,37} The experiments described here attempt to establish the order of Cu(I) binding to the target domain and the order of Cu(I) release from the active site of the donor. Also, this work suggests that a potential four-coordinate Cu(I) intermediate during ion transfer is energetically unfavorable.

In order to address these issues, the three Cu(I) cluster models were subjected to several calculations in Gaussian 03 so that free energy differences between the three models could be ascertained. In another experiment, the MM parameters defined for the system were used again in three kinds of free energy calculations using the AMBER suite. The three complete, solvated protein systems were subjected to thermodynamic integration calculations using single topology mutations, free energy perturbation calculations, and potential of mean force calculations.

Quantum Energy Calculations on the Small Cu(I) Thiolate Cluster Models

One difficulty in performing free energy calculations on the model system was the difference in number of atoms between different Cu(I) coordination states. For instance, a two-coordinate Cu(I) model cluster ($\text{Cu(I)[HSCH}_3\text{]}_2[\text{SCH}_3^-]_2^{-1}$) comprises twenty-three atoms, while the three-coordinate model ($\text{Cu(I)[HSCH}_3\text{][SCH}_3^-]_3^{-2}$) contains twenty-two atoms. Because of the difference in the number of atoms, a direct free energy comparison between the two models cannot be made. However, a comparison between models was able to be made by simulating an isodesmic reaction for protonation of a water molecule. An isodesmic reaction for the conversion of the two-coordinate model to the four-coordinate model with the transfer of two protons from the unbound methylthiols from the two-coordinate structure to two water molecules is shown below.



Single point energies were calculated for the gas-phase-optimized structures in both the gas-phase and implicitly solvated phase. In this manner, the $\Delta E_{\text{solvation}}$ could be calculated for each species in the reaction as well. The solvation correction to the gas-phase is important, especially for the charged species. Figure 3-6 shows the reaction for Cu(I) becoming four-coordinate from the three-coordinate state and the energy differences between products and reactants as well as $\Delta E_{\text{solvation}}$ for each molecule.

Figure 3-7 shows the reaction profile of changing the coordination of Cu(I) in the model systems. As expected, the addition of implicit solvent around the model mitigates the instability of the charged molecules in the gas phase. This is shown as the highly charged four-coordinate system as a much more favorable ΔE_{solv} than the two-coordinate

system. The figure illustrates the notion that the four-coordinate state is not energetically favorable, as it is 65.440 kcal/mol higher in energy than the three-coordinate model in the solvated state. Table 3-7 lists the energy differences between the different models in both the gas and solvent phases. The energy difference between the two-coordinate and three-coordinate models is lower by comparison at 48.531 kcal/mol. It should be noted that while two different three-coordinate structures can exist in the protein system, they are indistinguishable in the QM model. So while this experiment showed some relative energies between the three- and four-coordinate models, full protein simulations were needed to compare the energies of the two different three-coordinate states to each other.

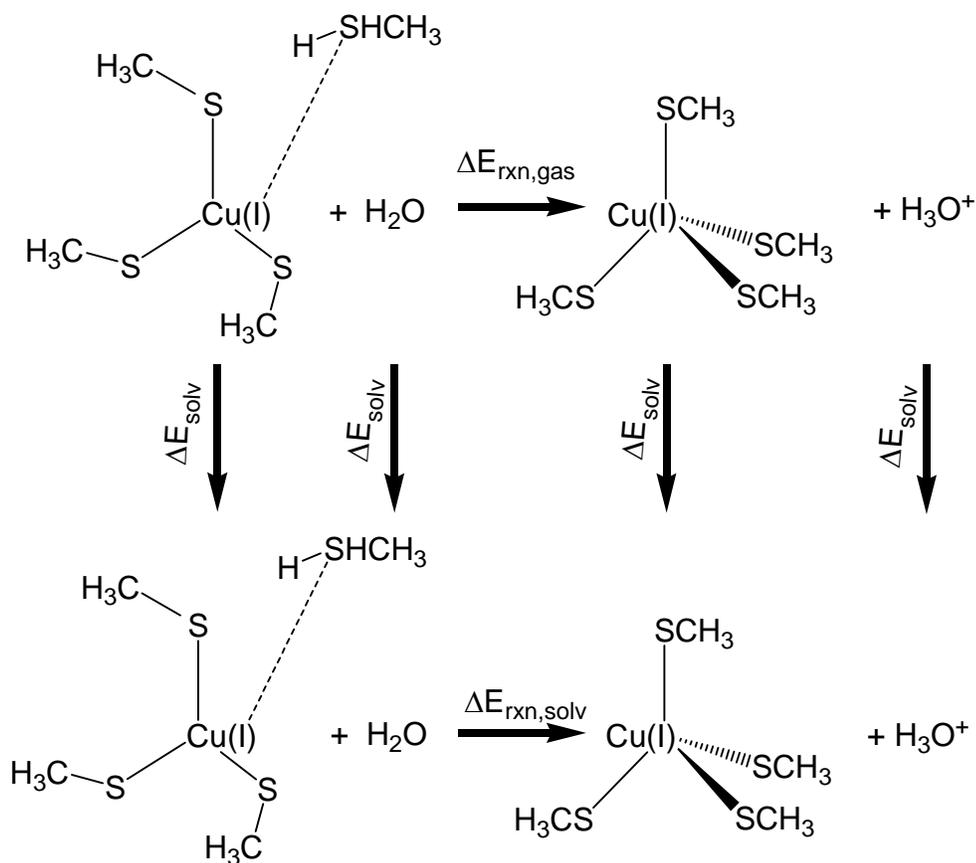


Figure 3-6. The isodesmic reaction and solvation of three-coordinate Cu(I) to four-coordinate in the model system.

Table 3-7. Reaction energies and energies of solvation for the model systems.

	Gas	Solvent
ΔE_{rxn} 3coord to 4coord	581	65.4
ΔE_{rxn} 2coord to 3coord	244	48.5
ΔE_{solv} 4coord + 2H ₃ O ⁺		-770
ΔE_{solv} 3coord + H ₂ O + H ₃ O ⁺		-254
ΔE_{solv} 2coord + 2H ₂ O		-59.1
Values in kcal/mol		

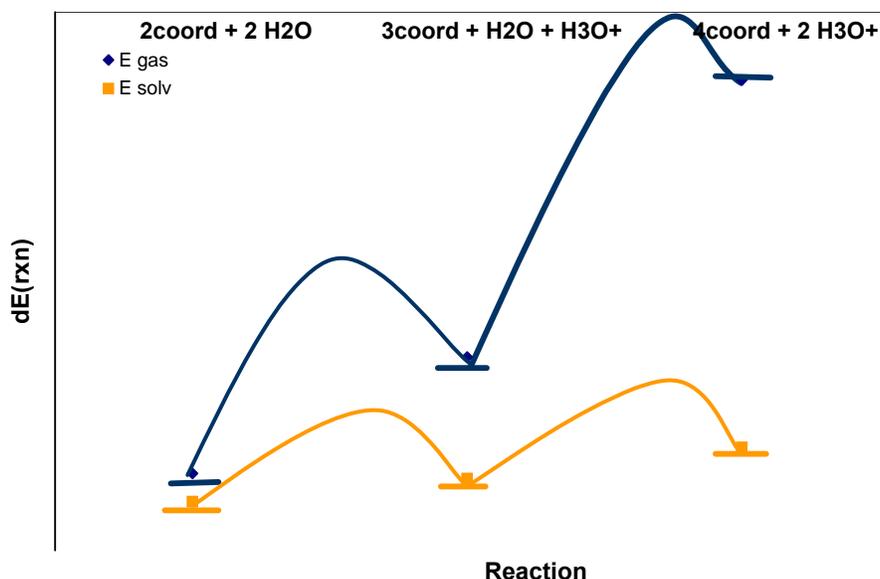


Figure 3-7. The relative energies of the species in the isodesmic reactions of the model systems in the gas phase (top) and in implicit solvent.

Free Energy Calculations on HAH1

Having developed the force field for the Cu(I) binding active site of HAH1, the challenge in setting up the free energy calculations lay in choosing which free energy method to apply and ensuring that the number of atoms in each simulation was the same for each different Cu(I) coordination state. The thermodynamic integration functionality of the *sander* program in AMBER 8 was chosen. The purpose of this study was to identify an energetically favorable route for Cu(I) transfer between two three-coordinate high affinity Cu(I) binding sites. From the two-coordinate state, there are two possibilities

for Cu(I) transfer. Earlier, the difference of the two Cys residues in the Cu(I) binding site was discussed. Cys 12 is more solvent exposed in the dimer state, while Cys 15 is along the interface between the donor and target Cu(I) binding domains and is not solvent-exposed. One possibility for Cu(I) transfer is for Cys 12 of the target to bind to the incoming Cu(I) ion first. The other option is for Cys 15 of the target to bind Cu(I) as the two domains come into close proximity.

The solvated, MD-equilibrated three-coordinate HAH1 dimer was initially used to create a starting point for the free energy calculations. This active site featured three Cu(I)-S bonds of approximately 2.3 Å along with a fourth longer distance Cu(I)---SH interaction to an unbound Cys residue at nearly 5.0 Å from the Cu(I) ion. In the initial state, both active site Cys residues of the donor domain bound Cu(I) while only one target Cys bound the Cu(I) ion. This structure was mutated in a stepwise fashion to the final structure which used both target Cys residues to bind Cu(I) while only one donor Cys continued to coordinate the metal ion. Figure 3-8 shows the proposed scheme for Cu(I) transfer. State 0 is the initial state and State 1 is the final state. Using a simple model system, such as that in the quantum work described earlier, there is no difference between State 0 and State 1 as shown below. However, when the entire solvated protein is simulated, conformational and solvation differences between the two states result in a free energy change as Cu(I) is transferred between the two active sites.

Although a simulation of the scheme in Figure 3-8 would yield a reliable value for the free energy difference between the two states, such a method is not currently applicable. The dual-topology free energy perturbation method was implemented in the *gibbs* program within AMBER. This method allows for the existence of dummy atoms in

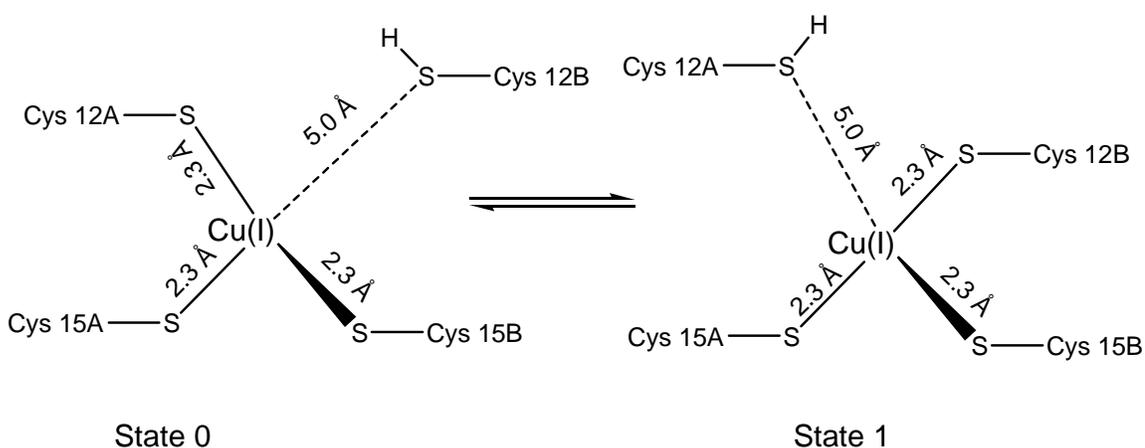


Figure 3-8. A scheme for Cu(I) transfer.

both the initial and final states of the simulation. A substantial attempt was made to perform a dual-topology calculation on the scheme presented above, but the simulation failed.

Since dual-topology could not be used to obtain the desired information, a single-topology approach was undertaken. Unlike a dual-topology approach, TI is not a continuous blending of initial and final reaction states as the simulation proceeds. Instead, TI calculations blend the reactant and product states in a series of discrete windows. These experiments used twelve windows to describe the reaction from start to finish. The first window would simulate a model whose character was very similar to the reactant, the sixth window would simulate a species whose character was nearly an equal blend of the force fields of the product and reactant, and the final window would simulate a system nearly identical to the product. Using TI in sander, it is possible to define an initial system with dummy atoms and mutate it to a structure without dummy atoms. In order to compare two different pathways for Cu(I) transfer, this experiment combined a set of two TI calculations. One calculation took the State 0 structure shown above and mutated it into a four-coordinate structure as defined in the previous section. This

simulation represented initial binding by Cys 15 of the target monomer. A similar simulation mutated State 1 from Figure 3-8 into the same four-coordinate structure. This represented initial binding by Cys 12 of the target domain. Since the two TI simulations shared the same endpoint, the energies of the two reactions could be compared. Figure 3-9 depicts the two reaction paths followed in the TI simulations.

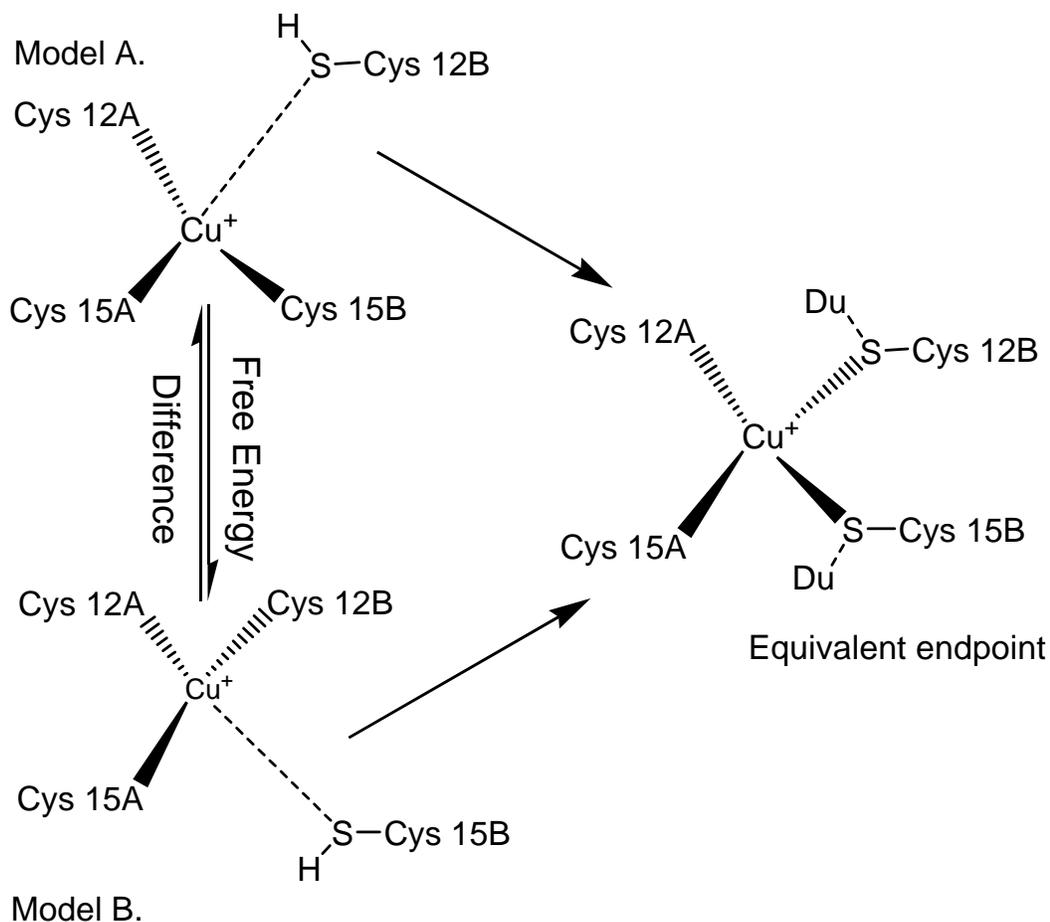


Figure 3-9. Two separate TI simulations were performed to compare two different Cu(I) transfer pathways from HAH1 to MNK4. The two reactions shown have different starting points, but the same endpoint. The top reaction is referred to as “Reaction 1” and the bottom is “Reaction 2” in the discussion that follows.

Each state contains the same number of atoms, but some atoms may have different atom types. Dummy atoms are used to as placeholders for the hydrogens that “appear” and “disappear” during the simulation. Dummy atoms have the same mass as hydrogen

atoms, but no charge. In each reaction, Cu(I) mutates from a three-coordinate state to a four-coordinate state, so its atom type must be changed. Initially, Cu(I) is defined as atom type PP, with bonds to S_{12A} (atom type SA), S_{15A} (SB), and S_{15B} (SD). A weak interaction is defined between PP and S_{12B} (SC). As the simulation progresses, the ion is mutated to type PQ which is bonded to SA, SB, SD, and S_{12B} (SF). Other atoms that change atom type are the sulfur atoms of Cys 12B and Cys 15B. In the top reaction of Figure 3-9, sulfur 12B is assigned atom type SC. It is bound to C_β from Cys 12B and a hydrogen of atom type HS (the default AMBER atom type for hydrogen bonded to S in Cys). SC also shares a weak interaction with Cu(I). The same atom is changed during the simulation to type SF which is bound to a dummy atom, C_β of Cys 12B, and Cu(I). The HS atom in the reactant becomes a dummy atom bound to sulfur SF in the product. SA, SB, and SC do not change atom type in the top reaction and are all considered to be equivalent. Bond angles between SF-PQ-SX (X=A,B, or D) in the product are included in the SF and PQ force fields.

The mutation scheme in the lower reaction is similar. In reaction two of Figure 3-9, the metal ion begins as atom type PP, and is bound to S_{12A} (SA), S_{15A} (SB), and S_{12B} (SC) and shares a weak interaction with S_{15B} (SD). Sulfur of Cys 15B is initially identified as atom type SD which is bound to a hydrogen (HS), C_β from Cys 15B, and has a weak interaction with the metal center. SD is mutated into atom type SE, which bonds a dummy atom and Cu(I) which is labeled as atom type PQ in the product of reaction 2. PQ in the lower reaction binds SA, SB, SC, and SE. In this reaction, SA, SB, and SC are treated equivalently and are not mutated in any manner.

Even though no bond exists between Cys 12B and Cu(I) at the start the reaction in Figure 3-9, a bond must still be defined. This is to satisfy the fact that bonds cannot be created nor destroyed in an MD simulation. In essence, the 5.0 Å Cu(I)-S_{12B} bond with a very weak arbitrary force constant in the reactant is being mutated into a 2.3 Å bond in the product with a well defined force constant. The same thing occurs over the course of the simulation of the lower reaction for the Cu(I)-S_{15B} bond. This treatment allows one bond to “form” with the target over the course of the simulation even though no actual bond exists in the initial state.

Tables 3-8, 3-9, and 3-10 list the atom types and force field parameters specified for the reactions shown in Figure 3-9. The parameters were derived from the three- and four-coordinate solvated and equilibrated structures from the long timescale MD simulations outlined above. Table 3-11 shows the atomic charges for the atoms in the two reactions of Figure 3-9. The charges on the ligating Cys residues can be compared to the default AMBER charges for Cys given in Table 3-4. For brevity, tables referring to atom type SX mean that the parameter is the same for all sulfur atoms within the active site. For example, SX-C_β refers to any bond between sulfur in the active site and C_β of the ligating Cys residue. Likewise, in Table 3-10, in angles referring to SX-PP SC, SX includes any S in that active site that is not SC. Some bond parameters, such as those for HS-SX and C_β-SX were adapted from the parm94.dat library instead of from QM calculations on the model clusters as were some bond angle parameters such as C_β-S-HS.

The TI calculations were performed on both gas-phase and explicitly solvated proteins. This was done in order to determine the solvation energy of the protein and any differences in active site geometry that may be caused by solvent. There were some

differences between the solvated and gas-phase simulations. No periodic boundary conditions were applied to the gas-phase system, which also was given a high cutoff (> 20 Å) for nonbonding interactions. The solvated system employed periodic boundary conditions, and the nonbonding cutoff was kept at the default value of 8.0 Å. The simulation of the solvated system was performed at constant pressure, while the temperature scaling was set for constant energy dynamics. For the solvated system using constant pressure dynamics, anisotropic pressure scaling was used in conjunction with the TIP3P water box.

Table 3-8. Atoms, atom types, atomic masses and van der Waals parameters used in TI simulations of Cu(I)-bound HAH1.

Atom	Atom type	Mass (au)	van der Waals radius (Å)	Well-depth (kcal)
Reaction 1				
Cu (reactant)	PP	63.55	2.40	0.05
Cu (product)	PQ	63.55	2.40	0.05
S12A	SA	32.06	2.00	0.25
S15A	SB	32.06	2.00	0.25
S12B (reactant)	SC	32.06	2.00	0.25
S12B (product)	SF	32.06	2.00	0.25
S15B	SD	32.06	2.00	0.25
HS 12B (reactant)	HS	1.008	0.60	0.015
HS 12B (product)	DU	1.00	0.00	0.00
Reaction 2				
Cu (reactant)	PP	63.55	2.40	0.05
Cu (product)	PQ	63.55	2.40	0.05
S12A	SA	32.06	2.00	0.25
S15A	SB	32.06	2.00	0.25
S12B	SC	32.06	2.00	0.25
S15B (reactant)	SD	32.06	2.00	0.25
S15B (product)	SE	32.06	2.00	0.25
HS 15B (reactant)	HS	1.008	0.60	0.015
HS 15B (product)	DU	1.00	0.00	0.00

Table 3-9. Bond length parameters for the reactions used in TI calculations of Cu(I)-bound HAH1.

Bond	k_{bond} (kcal/mol \AA^2)	r_0 (\AA)
Reaction 1		
PP-SA	60.000	2.190
PP-SB	60.000	2.190
PP-SC	0.001	5.000
PP-SD	60.000	2.190
HS-SC	274.000	1.336
C $_{\beta}$ -SX	219.354	1.849
PQ-SX	60.000	2.190
DU-SF	274.000	1.336
Reaction 2		
PP-SA	60.000	2.190
PP-SB	60.000	2.190
PP-SC	60.000	2.190
PP-SD	0.001	5.000
HS-SD	274.000	1.336
C $_{\beta}$ -SX	219.354	1.849
PQ-SX	60.000	2.190
DU-SE	274.000	1.336

The TI calculations do not encompass all of the contributions to the free energy change of the reaction. Because bonds cannot be broken or formed in MD simulations, weak interactions were described in places where bonds would be forming over the course of the simulations. This presents a problem in terms of how AMBER deals with bonding and nonbonding interactions. In AMBER, when two atoms share a bond, they are excluded from each other's nonbonding interactions. In our three-coordinate model, Cu(I) is supposed to be bonded to only three Cys ligands. However, since the fourth S had to be bonded to Cu with a weak interaction, its nonbonding interactions were being neglected. In reality, no bond exists between Cu and a fourth Cys ligand. To compensate for this, the topology files generated by LeAP had to be modified for the TI calculations so that the weakly bound S would be removed from the nonbonding exclusions of Cu and

its neighboring atoms. This allowed for the metal ion and the other Cys ligands in the active site to have nonbonding interactions with the unbound Cys.

Table 3-10. Bond angle parameters for the reactions used in TI calculations of Cu(I)-bound HAH1.

Angle	k_t (kcal/ mol rad ²)	θ_0 (deg)
Reaction 1		
SA-PP-SB	50.000	109.50
SA-PP-SD	50.000	109.50
SB-PP-SD	50.000	109.50
C _β -SX-PP	93.700	109.50
SX-PP-SC	0.001	109.50
HS-C _β -SC	0.001	109.50
H _β -C _β -SC	100.000	109.50
C _β -SC-PP	0.001	109.50
C _β -SC-HS	43.000	96.00
SX-PQ-SX	50.000	109.50
C _β -SX-PQ	93.700	109.50
H _β -C _β -SF	100.000	109.50
C _α -C _β -SX	50.000	109.50
C _β -SF-DU	43.000	109.50
PQ-SF-DU	50.000	109.50
Reaction 2		
SA-PP-SB	50.000	109.50
SA-PP-SC	50.000	109.50
SB-PP-SC	50.000	109.50
C _β -SX-PP	93.700	109.50
SX-PP-SD	0.001	109.50
HS-C _β -SD	0.001	109.50
H _β -C _β -SD	100.000	109.50
C _β -SD-PP	0.001	109.50
C _β -SD-HS	43.000	96.00
SX-PQ-SX	50.000	109.50
C _β -SX-PQ	93.700	109.50
H _β -C _β -SE	100.000	109.50
C _α -C _β -SX	50.000	109.50
C _β -SE-DU	43.000	109.50
PQ-SE-DU	50.000	109.50

While this approach solved one problem, it created another. TI simulations can only read in one topology file for each simulation. Therefore, once the fourth Cu(I)-S

bond formed by the end of the simulation, the atoms were still feeling the nonbonding interactions of the fourth Cys. As a correction to the TI calculation, the effects of forming that bond had to be determined. This was done using free energy perturbation. In separate simulations, the products of the two reactions in Figure 3-9 were used as the starting points for a FEP calculation. The perturbation would be the introduction of the vdW exclusions that were removed in the TI calculation.

Table 3-11. RESP charges used for TI calculations on Cu(I)-bound HAH1.

Atom type	RESP charge			
	Reaction 1		Reaction 2	
	Reactant	Product	Reactant	Product
PP	0.6483	n/a	0.6483	
PQ	n/a	1.3670		1.3670
SA	-0.8682	-1.0448	-0.8682	-1.0448
SB	-0.8682	-1.0448	-0.8682	-1.0448
SC	-0.8485	n/a	-0.8682	-1.0448
SF	n/a	-1.0448	n/a	n/a
SD	-0.8682	-1.0448	-0.8485	n/a
SE	n/a	n/a	n/a	-1.0448
HS	0.5470	n/a	0.5470	n/a
DU	n/a	0.0000	n/a	0.0000
N	-0.4157	-0.4157	-0.4157	-0.4157
H	0.2719	0.2719	0.2719	0.2719
C _α	-0.0351	-0.0351	-0.0351	-0.0351
H _α	0.0508	0.0508	0.0508	0.0508
C _β	0.0168	0.1011	0.0168	0.1011
H _β	0.0053	-0.04934	0.0053	-0.0493
C	0.5973	0.5973	0.5973	0.5973
O	-0.5679	-0.5679	-0.5679	-0.5679

Figure 3-10 displays the FEP scheme as a correction to the TI calculations. A four-coordinate model (the product of the TI simulation) with vdW exclusions removed between the newly bound Cys and the rest of the active site was simulated using the with the vdW exclusions intact. The FEP calculation was a trajectory analysis of the exclusions-removed structure using the exclusions-present Hamiltonian and only took

one step. The free energy difference between the initial and final states of the FEP simulations equaled the contribution to the total free energy of nonbonding interactions becoming bonding interactions as the new Cu(I)-S bond formed in the TI simulation.

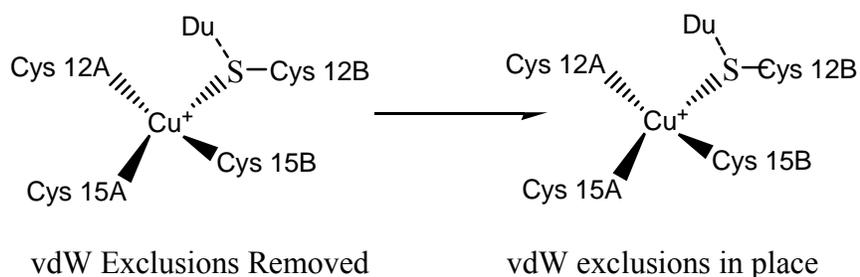


Figure 3-10. FEP vdW correction to TI on HAH1. Evaluate trajectory of structure with vdW exclusions removed with the Hamiltonian with vdW exclusions intact.

The FEP simulations revealed another contribution to the total free energy change between the three- and four-coordinate proteins. While the Cu(I)-S bond lengths for the three initially bound Cys ligands remained unchanged throughout the TI and FEP simulations, the new Cu(I)-S bond did not reach the correct length. This was due to the fact that AMBER did not allow the new S to come any closer than about 2.8 Å to the Cu(I) ion while the bonding interactions were turned off. In a sense, the penalty for removing the vdW exclusions was not only the omission of bonding interactions once the new bond had been formed, but also that the new bond was too long. The normal Cu(I)-S bond length was around 2.2 Å with a force constant of 60.00 kcal/mol·Å, but the newly formed Cu(I)-S bond was 2.8 Å with the same force constant.

Another correction to the TI simulation had to be made in the form of “reeling in” the newly-bound S to the metal center. The energy profile of shortening the bond length could be generated by a potential of mean force calculation during which the products of the reactions listed in Figure 3-9 (with vdW exclusions in place) would again be used as starting structures. The four-coordinate structures featured three Cu(I)-S bonds of the

correct length, and one Cu(I)-S bond that was too long. The fourth S bond would be contracted from 2.8 Å to 2.0 Å over the course of twenty-three windows in the PMF simulation. Figure 3-11 shows the reaction scheme for the PMF experiment, and Figure 3-12 shows the energy profile of shortening the final Cu(I)-S bond. A steep harmonic potential was induced upon the long Cu(I)-S bond with a minimum at 2.1 Å. The energy difference between the initial bond length and the minimum energy bond length on the PMF curve served as the third and final contribution to the free energy change of the reactions in Figure 3-9. The data from the PMF experiments were connected using weighted histogram analysis with the WHAM software.^{24,25}

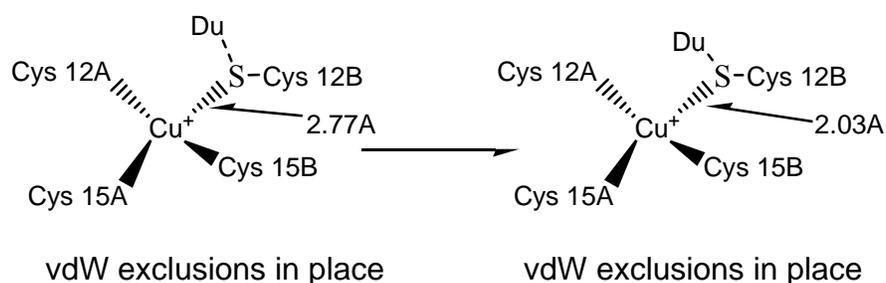


Figure 3-11. Bond length correction to FEP calculations by PMF: contract Cu(I)-S bond length from ~2.8 Å to ~2.0 Å by PMF analysis of twenty-three windows.

The overall free energy change from three- to four-coordinate Cu(I) in HAH1 is the sum of the TI mutation, the FEP trajectory analysis for vdW interaction correction, and the PMF for bond length correction. Table 3-12 lists the free energy changes for the TI reaction shown above for both gas-phase and aqueous systems. As shown, the addition of solvent lowers the energy barrier of binding the fourth S to Cu(I). Recalling Figure 3-9, the endpoints of each reaction are equivalent. So the free energy difference between the two different three-coordinate reactants can be determined by taking the difference between the total free energy differences of their respective reactions.

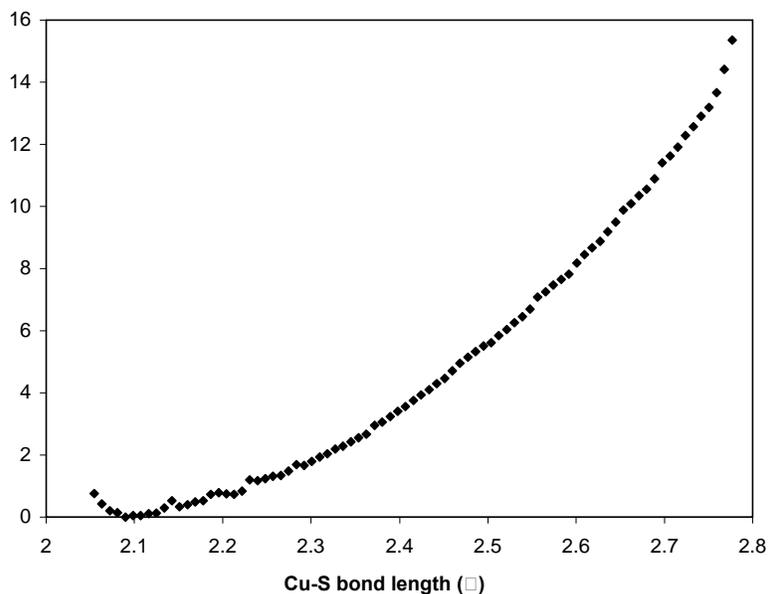


Figure 3-12. PMF curve of solvated HAH1 showing minimum energy Cu(I)-S_{12B} bond length near 2.1 Å for the bonding of Cys 12B to Cu(I).

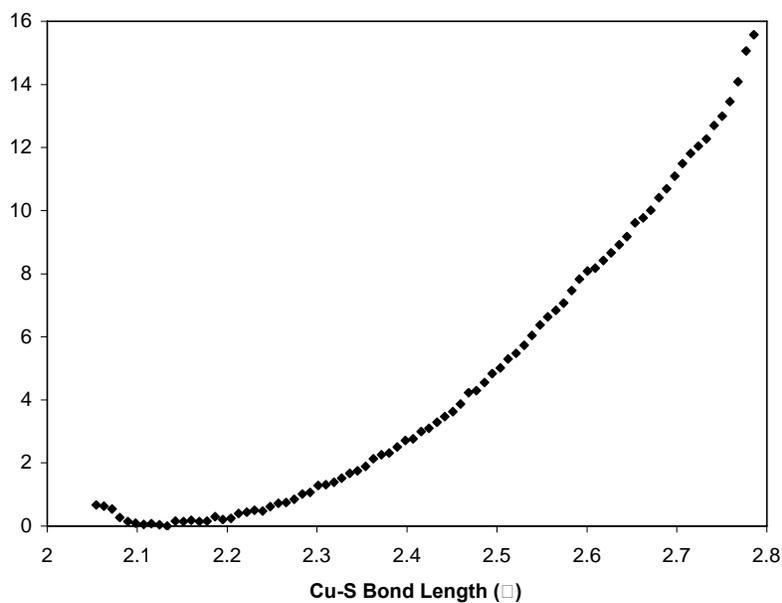


Figure 3-13. PMF curve for the binding of Cys 15B to Cu(I) in solvated HAH1, showing a minimum energy bond-length of just over 2.1 Å for the Cu(I)-S_{15B} bond.

Table 3-12 shows the free energy change of the reactions displayed in Figure 3-9, and Figure 3-14 plots the free energy difference between the two different three-

coordinate states in the explicitly solvated protein. These values show that the Model A structure of Cys 15B of the target monomer binding Cu first is energetically favorable to Cys 12B binding Cu first by 24.7 kcal/mol.

Table 3-12. Free energy changes for TI calculations on the reactions shown in Figure 3-9.

TI ΔG	
Model A: Cys 15B binding 1st	
Gas	224.2
Solvent	177.7
Model B: Cys 12B binding 1st	
Gas	213.9
Solvent	153.0

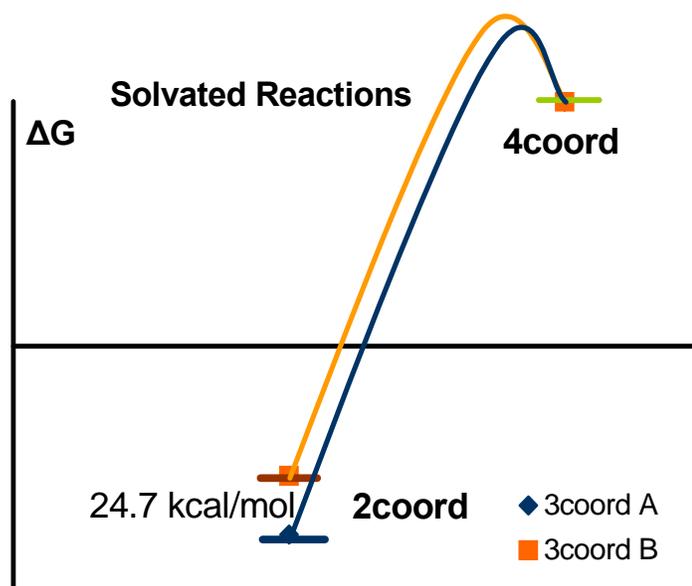


Figure 3-14. The free energy difference by thermodynamic integration between the different three-coordinate Cu(I)-bound HAH1 dimers.

Table 3-13. The free energy difference of changing the coordination environment of Cu(I) in HAH1.

	Cys 12B unbound	Cys 15B unbound	ΔG
Gas	255.5	252.8	-2.7
Solvent	219.5	193.8	-25.7
Solvent effect			-23.0

Values are in kcal/mol.

Conclusions

The results from the QM and MM studies on the model systems and the HAH1 dimer can be interpreted to suggest an energetically favorable order of Cu(I) transport between the active site of a donor HAH1 monomer and the active site of the fourth domain of the Cu(I) receptor MNK. The QM calculations done in the first part of this experiment created a foundation for the description of the Cu(I) binding site in HAH1. Further QM work detailed the thermodynamics of Cu(I) thiolate clusters as models of the active sites of the MT/CXXC family of Cu(I)-binding metalloproteins. The first part of the molecular dynamics study was to create a force field to describe the atoms involved in Cu(I) binding in HAH1 based on the QM calculations. Then, MD simulations were performed with the new force field. Analysis of these simulations showed the accuracy and reliability of the new force field parameters. The final stage of the experiment was an investigation of the free energy of Cu(I) transport between two metal binding domains.

The HAH1 dimer was used as a model for Cu(I) transfer from the active site of a HAH1 monomer to the fourth domain active site of the Wilson's disease protein. In this model, Cys 12A and Cys 15A of the HAH1 dimer represent the donor active site, while residues Cys 12B and Cys 15B represent the metal binding site of MNK4 which are Cys 14 and Cys 17, respectively. Mechanistically, the free energy calculations suggest that when a Cu(I) is being transferred from the HAH1 binding site to the MNK4 site, that Cys 17 of the MNK protein fourth domain is energetically more favorable to bind the incoming Cu(I) before the more solvent-exposed Cys 14. Physically, this makes sense due to the fact that the solvent-exposed Cys 14 is farther away from the protein-protein interface than Cys 17 on the target domain and that solvent interactions would stabilize Cys 14 on the surface of the protein. At that point, Cys 12 of the donor domain would

start releasing Cu(I) as Cys 14 of MNK4 started to bind the ion. There is no evidence that a purely four-coordinate Cu(I) species exists during copper transport. This is supported by the QM results early in the study. Instead, it appears that the Cu(I) ion is nearly always three-coordinate as it is transferred between the two proteins. In the proposed transfer mechanism, Cys 15 of HAH1 is the last donor residue to release the copper ion. When copper transfer is complete, HAH1 is no longer bound to copper and the active site of MNK4 complexes the Cu(I) ion.

CHAPTER 4
ELECTRONIC STRUCTURE OF THE ACTIVE SITE OF AMINOPEPTIDASE
FROM *Aeromonas proteolytica*

AAP Introduction

Zinc-dependent peptidases such as bovine lens leucine aminopeptidase (bLAP), carboxypeptidase A, thermolysin, and the aminopeptidase from *Aeromonas proteolytica* (AAP) play important roles in tissue repair, carcinogenesis, protein maturation, cycle cell control, the regulation of hormone levels,^{56,57} and the degradation of DNA, RNA, phospholipids, and polypeptides.⁵⁸ Improper functioning of aminopeptidases has been linked to health issues including aging, cataracts, inflammation, cystic fibrosis, cancer, and leukemia.⁵⁶⁻⁶⁰ Despite the variety of cellular processes in which aminopeptidases are involved, not much was known about their exact functions or mode of action until recently. Peptidases such as carboxypeptidase A and thermolysin which utilize a sole Zn^{2+} ion for catalysis have been extensively studied and their modes of action are relatively well understood.^{56,58,60} Aminopeptidase from *Aeromonas proteolytica* is a dinuclear metallohydrolase which employs two Zn^{2+} ions to catalytically cleave the N-terminus of a polypeptide chain. Its small size (~ 32 kDa), high thermal stability, and functionality as a monomer^{61,62} led AAP to being one of the first peptidases to be isolated and characterized in detail.⁶³ Substituting the spectroscopically silent Zn^{2+} ions with Co^{2+} or Cu^{2+} allowed for further kinetic and mechanistic studies on the protein and did not adversely affect catalytic activity.^{63,64} In fact, some hyper-active species of AAP were created by these substitutions.⁶⁵

Native AAP (Figure 4-1) contains two Zn^{2+} ions in the active site, but can perform its function at 80% efficiency with only one Zn^{2+} present.⁶⁶ In fully functioning AAP, both cations are present and perform some catalytic function. The reason why some peptidases function in a mononuclear capacity while others require multiple ions for full efficiency is not yet understood.⁶⁶ The binding pocket in AAP has been shown to bind all N-terminal amino acids and can accommodate all penultimate residues except Glu and Pro. Being largely hydrophobic in nature,⁶⁷ the active site preferentially binds hydrophobic residues with Leu being the most easily cleaved.⁶²

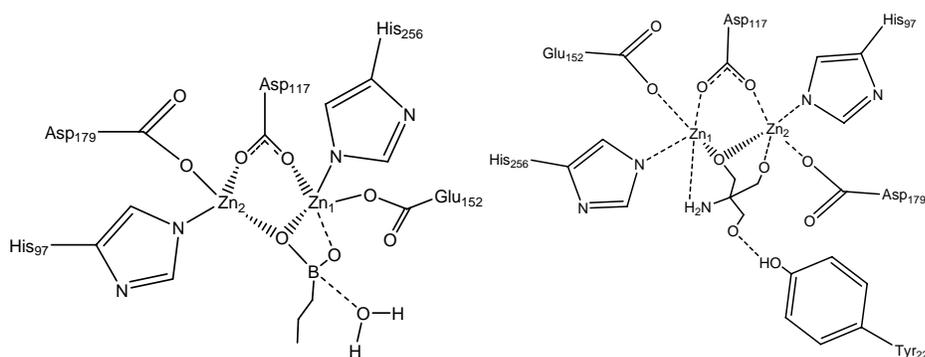


Figure 4-1. AAP active site inhibited by Tris (left) and BuBA . Investigation of the X-ray structures of these complexes shed light on substrate conformation and a potential mechanism for peptide hydrolysis in AAP. PDB ID 1LOK, 1CP6

The metal-binding pocket of AAP is characterized by several Asp, Glu, and His residues which coordinate the Zn cations. X-ray crystallographic studies on native AAP have predicted a tetrahedral (T_d) geometry for both cations when no substrate is present,⁶⁸ although in its closed-shell electronic state Zn^{2+} shows no preference for either octahedral (O_h) or T_d geometry.⁵⁷ Beyond the divalent cations, other catalytically important features of the binding site include the bridging water/hydroxide molecule and Glu151 each of which have potentially important roles in the proposed catalytic mechanism of AAP. In 1992, Chevrier *et al.* were the first to produce a high resolution (1.8 Å) crystal structure

of native AAP. This pioneering work not only showed that the active site was dinuclear, but it also identified the key first shell Zn-complexing residues Asp117, Asp179, Glu152, His256, and His97.⁶⁷ Further high resolution crystallographic studies on inhibitor-bound AAP carried out by several groups have since clarified the roles of second shell complexing residues such as Glu151, Tyr225, Ser228, Cys227, and Asp99, the importance of the bridging water/hydroxide and other water molecules in the active site, active site coordination geometries upon substrate binding, and have led to the proposal of several catalytic mechanisms.^{62,66,69-71}

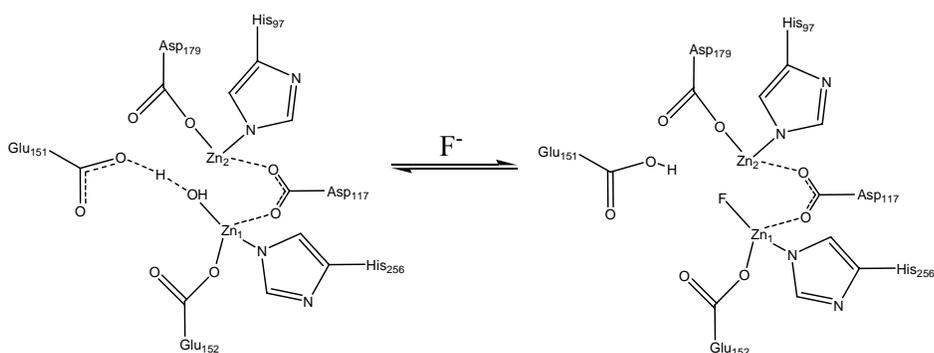


Figure 4-2. In fluoride inhibition studies of AAP, it was shown that a F^- ion displaces a terminal hydroxide group, deactivating the enzyme.

Several inhibition studies have been performed on this system to complement the crystallographic work. The twofold purpose of these studies has been to both characterize the nature of the inhibited protein and to investigate possible drug candidates for enzyme inhibition. Beyond the preference for hydrophobic residues in the binding cleft, potential substrates should have a free α -amino group in the *L*-configuration.⁵⁷ At present, several well-known peptide inhibitors have been shown to inhibit AAP. Potent inhibitors include *L*-leucinethiol, hydroxamates, α -hydroxyamides, and notably 1-butaneboronic acid (BuBA) and Tris.^{57,59,62,63,72-76} Inhibitor binding to both cations is not necessary for AAP inhibition, and X-ray structures of both the Tris- and BuBA-inhibited enzyme (Figure 4-

2) have revealed that the water/hydroxide bridge between the cations is broken.⁷² These data suggest that the μ -aqua bridge is broken to form a terminal hydroxyl- group at some point during peptide hydrolysis in order for the enzyme to function properly. This hypothesis is confirmed by fluoride inhibition studies of native AAP.^{57,77} A single fluoride ion binds to Zn₁ in the active site (Figure 4-3), displacing the terminal water/hydroxyl- group after substrate binding, and the reaction does not proceed. Inactivation only occurs after substrate binding, suggesting that a terminal hydroxyl- group is not present until the carbonyl oxygen of the activated scissile bond has bound to Zn₁ and peptide hydrolysis is underway. Interestingly, chloride ions do not inhibit AAP up to a 2 M concentration because they do not bind with sufficient strength to the cation in the active site.^{57,77} The highest resolution structure of AAP was obtained in 2002 by the Petsko lab.⁵⁹ The 1.2 Å structure of native AAP in Tris buffer reduced the amount of structural uncertainty due to side chain motion, determined the position of several hydrogen atoms in the protein, and clarified to some degree the geometry of the Tris-bound active site.⁵⁹ Of note is how the distances between the Zn ions and the bridging O atom change from the unbound native structure 1AMP and the Tris-bound structure 1LOK. In the unbound protein Zn₁-O and Zn₂-O distances are 2.29 Å and 2.25 Å, respectively. The Tris-bound active site reveals Zn₁-O and Zn₂-O distances of 1.95 Å and 2.21 Å. The change in Zn-O distances may be further evidence of the conversion of the bridging water/hydroxide group into a terminal moiety.

Both cations must be present in the active site in order for AAP to be fully efficient, and they must have a task to perform during peptide hydrolysis. Zn²⁺ has been shown to be a hard acid,⁷⁸ and the mono-zinc environment has been shown by Christianson and

Cox to reduce the pK_a of a single water molecule in bulk solvent from 15.7 to 9.0.^{59,79}

The pK_a of bound water in a di-zinc environment is expected to be much less than 9.0. It has been proposed that the two cations, each acting as a Lewis acid, perform separate but equally important functions in the reaction cycle. A common thread between several

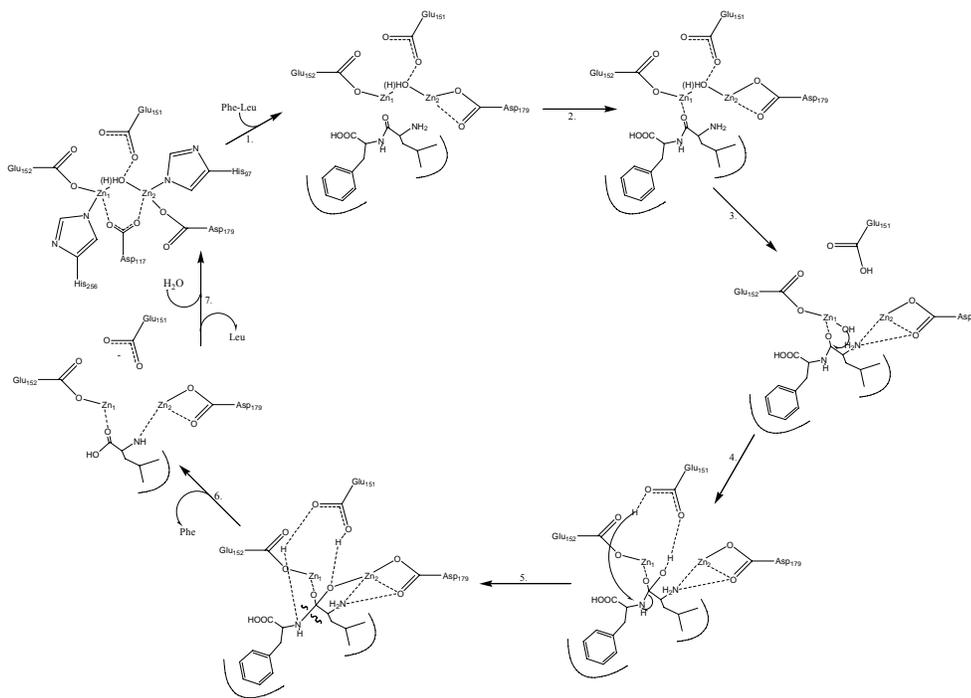


Figure 4-3. A proposed mechanism for AAP peptide hydrolysis showing proton transfer to Glu151, formation of a terminal hydroxyl- group, a gem-diolate intermediate, donation of a proton back to the leaving amino group, and reformation of the water/hydroxide bridge. Adapted from Petsko.⁵⁹

proposed mechanisms has been that the N-terminal amino group binds to Zn₂ and Zn₁ binds the carbonyl oxygen of the activated scissile bond. The mechanism proposed by Stamper *et al.* based on kinetic, crystallographic, and spectroscopic studies shows Zn₁ binding to the carbonyl group of the scissile bond, followed by the N-terminal amino group binding to Zn₂.^{57,66,71,77,80-82} Holz reasons that carbonyl binding occurs prior to amino binding as a result of inhibition studies of LeuSH on [CoCo(AAP)].^{57,73} The observed geometry of boron in the study of BuBA-inhibited AAP is further evidence of

this binding sequence.^{57,70} As stated earlier, other key players in AAP peptide hydrolysis are Glu151 and the bridging water/hydroxide group. In proposed mechanisms, a bridging or terminal OH⁻ would serve as a nucleophile and Glu151 would act as a general base.^{57,61,62,66,71,77}

Beyond the observations made from fluoride inhibition studies, there is further evidence of a terminal hydroxyl group. BuBA inhibition studies show that when a substrate is bound, the distance between Zn₂ and Asp117 and Asp179 is decreased to 3.0 Å from 3.4 Å in the native structure. The decreased distance allows for the formation of a strong hydrogen bond between Glu151 and His97 that does not exist in the unbound protein. Substrate binding also brings Asp99 closer to His97, creating yet another hydrogen bond. The proximity of the two negatively charged residues to Zn₂ along with the newly formed H-bonds effectively stabilizes the charge neutrality of Zn₂ and regulates its Lewis acidity. A sufficient decrease in the acidity of Zn₂ would facilitate the formation of a terminal water/hydroxide group on Zn₁.^{57,70,83} Chen *et al.*⁷⁷ suggested that Glu151 assists in the deprotonation of a terminal water molecule, resulting in a nucleophilic hydroxo- moiety, followed by attack by that group on the carbonyl oxygen of the scissile peptide bond, forming a *gem*-diolate intermediate characterized by two oxygens binding to Zn₁. The *gem*-diolate is stabilized through its interaction with both Zn ions.⁵⁹ At this point, the reaction proceeds toward completion with Glu151 donating a proton back to the penultimate amino group (now the N-terminus of the leaving group), which departs the binding cleft upon C-N bond cleavage, the rate-limiting step of peptide hydrolysis.⁷⁷ The final step is the reformation of the water bridge between Zn₁ and Zn₂ as

the active site returns to its native unbound conformation. Other publications have since supported the mechanism proposed by Chen *et al.*^{57,59,84}

Many mechanisms have been suggested for peptide hydrolysis by AAP, and there are some contentious points among them. Overall, assumptions are made about protonation states of the water bridge and Zn-binding residues, and the conformation of the substrate in the active site. Desmarais *et al.* contend that uncertainties in the reaction mechanism can not be clarified without more detailed knowledge of the electronic structure and protonation state of the metal ions, water molecules, and residues in the immediate active site.⁵⁹ Despite their exhaustive QM/MM study of the AAP peptide hydrolysis mechanism, Schürer *et al.* suggest that molecular dynamics simulations of the protein are needed to take accurate account of conformational movements of the protein and substrate.⁷¹ While Schürer *et al.* suggest that high level *ab initio* or DFT studies on the complete AAP active site would be prohibitively expensive, those experiments have been performed in this study. A series of those calculations have produced some data about the electronic structure and geometry of several active site protonation states.

Effects of 1st-Shell Mutations

Numerous inhibition, crystallographic, kinetic, and computational studies have been performed on the aminopeptidase from *Aeromonas proteolytica* (AAP) in order to gain a better understanding of the mechanism of the peptide hydrolysis reaction catalyzed by the enzyme.^{57,59,62,66,67,71,84} However, the research performed on AAP to this point has yet to answer key questions regarding the protonation state of the Zn-Zn bridging species in the native and active states of the enzyme, the role of Glu151 in the reaction, and the electronic structure of the dinuclear center. Crystallization of Tris-inhibited AAP and its structural characterization by XRC to a resolution of 1.2 Å yielded new information

about the side chain conformations of several residues in inhibited AAP as well as the positions of some hydrogens in the enzyme.⁵⁹ However, that study was not able to determine the nature of the Zn-Zn bridge in the protein. Schürer, Lanig, and Clark completed a detailed QM/MM study of the AAP peptide hydrolysis mechanism by determining relative energies of several possible intermediate and transition state species using AM1 and VAMP for the QM and MM regions, respectively.⁷¹

This computational work entails fully quantum geometric and energetic optimization of the AAP active site using B3LYP/6-31G* in Gaussian 03.⁸⁵ Here, data is presented from these studies pertaining to the electronic structure and coordination of the di-zinc-containing AAP active site. The active site model employed in these calculations is similar in nature to the one investigated by Schürer *et al.*, comprising the side chains of Asp117, Asp179, His256, His97, Glu152, Glu151, two Zn²⁺, the bridging species, and crystallographic water molecules within the active site. However, none of the structures of 1st-shell mutations include an inhibitor molecule or the second-shell residues Asp99, Cys227, Ser228, and Tyr225. The initial active site geometry was obtained from the 1.8 Å resolution crystal structure of AAP (Figure 1, PDB-ID 1AMP) obtained by Chevrier *et al.* in 1992.⁶⁷

This structure was modified by the removal of the backbone atoms of each residue except for the bridging atoms between Glu 151 and Glu152 and the addition of one or two protons to the bridging oxygen. Other structures were created by protonating Glu151 at the oxygen closest to the di-zinc bridge. The initial structures were used to generate B3LYP/3-21G* optimized geometries in Gaussian 03. Those models were in turn used as starting structures for the final B3LYP/6-31G* optimization. Single point

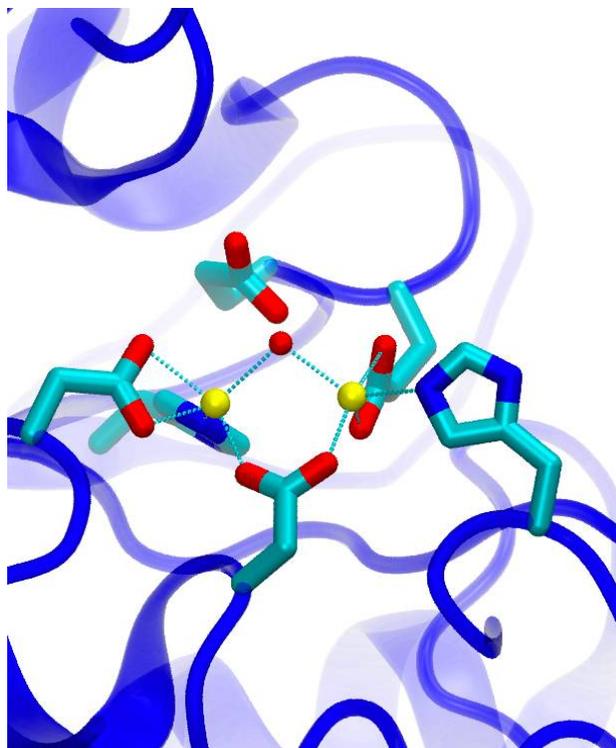


Figure 4-4. The general model for the QM work is the AAP active site from PDB structure 1AMP, the 1.8 Å resolution structure elucidated by Chevrier and Schalk.⁶⁷ Asp 117 is below the two Zn ions, with Zn₂ on the left and Zn₁ on the right. The residue at the top of the active site is Glu 151. Zn₂ is bound to His 97 and Asp 179, and Zn₁ is complexed with His 256 and Glu 152.

energy calculations at a higher level of theory, such as MP2, have not yet been attempted because the computational expense of such an experiment would be too high, and the resulting energies from the geometry optimizations performed here are sufficiently accurate for future mechanistic studies. Ultimately, the goal of this work is threefold: to investigate the different protonation states of the water/hydroxide bridge, to measure C-O bonds in the Zn-coordinating carboxylate residues, and to gauge the importance of Glu151 as a proton acceptor in the initial stages of AAP peptide hydrolysis.

By performing calculations on an array of protonation states, the relative energies between possible intermediates of the hydrolysis reaction were able to be determined, namely in a potential initial proton transfer from the water bridge to Glu151. The first

system that was investigated was one where a water bridge exists between Zn₁ and Zn₂. This species was compared to an active site with a hydroxide bridge and a protonated Glu151, with both models containing 78 atoms. The relative energy of the optimized structure of the hydroxide-bridged state shows that it is energetically favored over the water-bridged state by more than 4.10 kcal/mol. However, upon inspection of the optimized water-bridged structure, it was seen that one hydrogen from the water bridge transfers to Glu151. Another feature of the optimized water-bridged structure is the formation of interactions between E152 and H256 and between Zn₁ and crystallographic water that was retained in the active site. The optimized structure of the hydroxide-bridged model with E151 initially protonated reveals protonation of D179. D179 gains a proton from one of the water molecules retained in the active site, while the hydroxide ion formed by that deprotonation interacts with H256 on the other side of the active site. In the end, the energy difference between the two 78-atom models may be attributed to the different interactions and conformations that form during the optimization.

The next study compared a model with an initial hydroxide bridge and unprotonated Glu151 to an oxygen-bridged active site with Glu151 being protonated, with each model containing 77 atoms. The hydroxide-bridged model is 4.62 kcal/mol more favorable. Upon comparison of multiple species, the structures with an OH⁻ bridge are lower in energy than either the O²⁻-bridged model or the H₂O-bridged model. This trend was also observed during a simple single point energy comparison between different protonation states of the native crystal structure without geometry optimization. This may suggest that favorable intermediates in the reaction mechanism may all have bridged or terminal OH⁻ species as opposed to O²⁻ or H₂O bridges.

These data generally support the previous proposal that a high-energy water-bridged active site would initially donate one of its hydrogens to Glu151 in order to produce a nucleophilic hydroxide-bridging species in an exothermic process.⁷⁷ Further interpretation of these results suggests that an initial hydroxide bridge would not donate its hydrogen to Glu151. In that instance, Glu151 would not have a proton that it could donate to the N-terminus of a polypeptide chain within the active site as proposed. Both systems reveal the stability of a hydroxide bridge over an oxyl- or water-bridge between Zn₁ and Zn₂. In all cases, the crystallographic water molecules in the active site work together with the bridging species and the surrounding carboxylate residues to establish a robust hydrogen-bonding network within the active site. A more detailed discussion of the optimized geometries of several models is included below.

Table 4-1. Electrons in the side chains of Asp117 and Asp179 are equally delocalized over the carboxylic acid region, while Glu151 and Glu152 side chains contain one C-O bond with more electron density than the other.

	Asp117		Asp179		Glu151		Glu152	
	C _β -O _{γ1}	C _β -O _{γ2}	C _β -O _{γ1}	C _β -O _{γ2}	C _γ -O _{δ1}	C _γ -O _{δ2}	C _γ -O _{δ1}	C _γ -O _{δ2}
H ₂ O bridge	1.28	1.26	1.26	1.28	1.24	1.31	1.36	1.22
OH ⁻ bridge	1.27	1.26	1.26	1.27	1.28	1.32	1.30	1.24
O ²⁻ bridge	1.27	1.27	1.25	1.28	1.22	1.33	1.28	1.25
GluH + O ²⁻ bridge	1.26	1.27	1.24	1.30	1.22	1.33	1.28	1.25
GluH + OH ⁻ bridge	1.27	1.27	1.24	1.30	1.23	1.33	1.29	1.25
X-ray	1.22	1.24	1.22	1.22	1.24	1.23	1.23	1.22
Values in Å								

C-O bond lengths for Asp117, Asp179, Glu151, and Glu152 are listed in Table 4-1. It is clear that both C-O bonds equally share the electrons in the carboxylic regions of the aspartic acid residues as the bond lengths are nearly identical. The acidic regions of the Glu residues do not share this feature. Instead, one C-O bond is clearly higher in electron

density, while the other C-O bond is comparatively longer. These data suggest that the oxygens in the Asp117 and Asp179 are coordinating equally into the metallic center of the active site. On the other hand, only one oxygen of the side chains of Glu151 and Glu152 is coordinating with the zinc centers. In the case of Glu151, the side chain serves a role in the H-bonding network that exists throughout the active site.

Along with the protonation state of the bridging species, the metal-binding carboxylic amino acids Asp117, Asp179, and Glu152 are interesting research subjects. An analysis of C-O bond lengths and O-Zn distances help describe bond order and electron density and metal ion coordination, respectively. Measurement of the C-O distances in the carboxylic acid side chains of the Zn-coordinating residues yields information about the localization of electrons within the carboxyl regions of the coordinating residues. Table 4-2 lists Zn-Zn, Zn-O, and Zn-Asp117 distances for the structures shown in Figures 4-5 and 4-6. Generally, Zn-Zn distances are around 3.3 Å, which is shorter than the inter-zinc distance in the 1AMP crystal structure. The one exception is for structure 4-6f in which a terminal peroxo- group exists and the two zinc ions are separated by more than 4.0 Å.

Table 4-2. Several distances are shown for Zn-Zn and Zn-O interactions for the structure shown below in Figures 4-6 and 4-7.

	Zn ₁ -Zn ₂	Zn ₁ -O _μ	Zn ₂ -O _μ	D117-Zn ₁	D117-Zn ₂	E151-O _μ
H ₂ O bridge	3.31	1.96	2.02	1.99	2.00	2.55
OH ⁻ bridge	3.28	1.96	1.97	1.96	2.00	3.28
O ²⁻ bridge	4.23	1.97	4.03	2.02	2.01	3.27
GluH + O ²⁻ bridge	3.37	2.05	1.99	2.01	1.98	2.71
GluH + OH ⁻ bridge	3.26	1.96	1.95	1.96	1.97	2.77
X-ray	3.47	2.25	2.29	2.05	2.01	3.30
Values in Å						

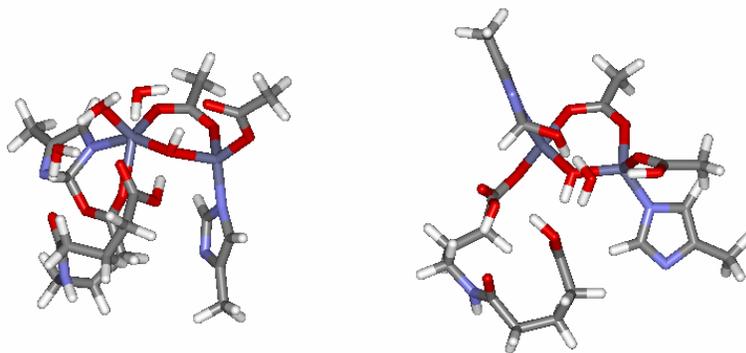


Figure 4-5. B3LYP/6-31G* optimized geometries of two models of the AAP active site. Asp 117 is shown in the upper-right of both pictures, binding each Zn. Zn₁ is the ion on the left and Zn₂ on the right in each structure. The structure on the left is from an originally water-bridged structure and Glu151 has gained a H, while the structure on the right started with a OH⁻ bridge and Asp 179 gains a H from a crystallographic water.

The geometry optimizations of several variations of this active site model are discussed here. Generally, the starting structures are the same in each case, and the first three models we investigated differ only by the protonation state of the bridging group while Glu151 is unprotonated (Figure 4-6). As shown in Figure 4-5a, an initial water bridge with an unprotonated Glu151 is optimized to a OH⁻ bridge as the initial water loses a hydrogen to nearby Glu151. The Zn-Zn distance decreases from the 3.47 Å shown in the original crystal structure to 3.31 Å in the optimized structure. The optimization of the second model (Figure 4-5b) is more complicated as the initial OH⁻ bridge and unprotonated Glu151 becomes a bridging O₂H peroxo- species and Glu151 and Glu152 are both protonated. It appears that a crystallographic water and the initial bridging OH⁻ donate one H each to Glu151 and Glu152.

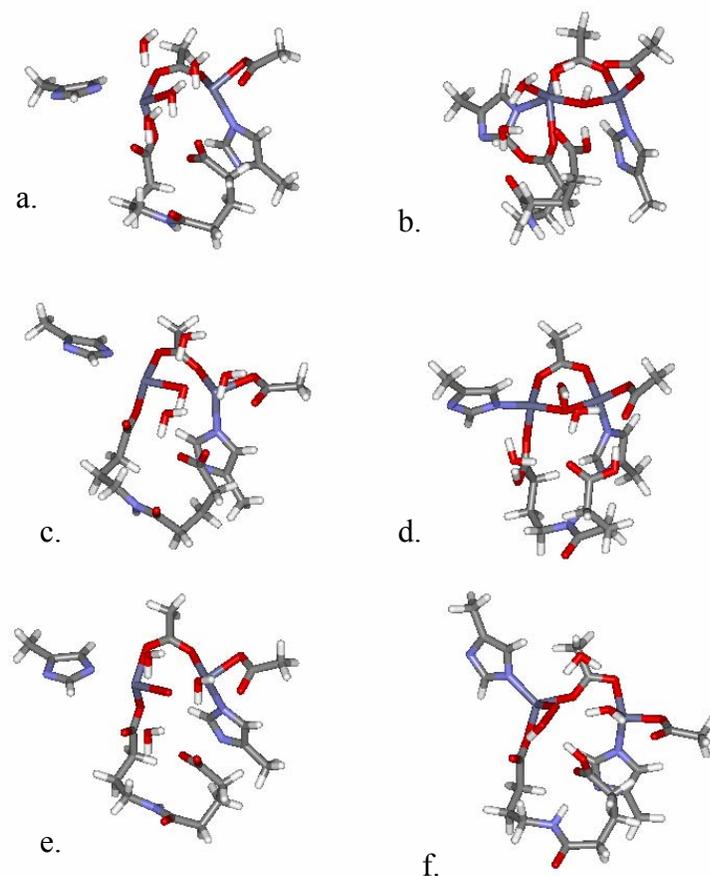


Figure 4-6. Starting structures (left) and B3LYP/6-31G* optimized geometries for different Zn-Zn bridging species within the active site of AAP. a) a water bridge b) a hydroxyl- bridge c) an oxyl- bridge.

This type of bridging group has not yet been discussed as a possibility in the proposed reaction schemes. However, the formation of the peroxy- group may be one consequence of not modeling an inhibitor into the active site. Not this optimized structure nor any other optimized structure containing a peroxy- bridge was shown to have the lowest relative HF energy to structures with a similar number of atoms. The last model of the first group (Figure 4-6b) contains a bridging O^{2-} ion. This optimization results in the formation of two terminal species. On Zn_1 , a terminal peroxy- group forms, similar in nature, but not geometry to the peroxy- group formed during the optimization of structure

4-5b. Then, one of the crystallographic water molecules becomes terminal to Zn_2 . Glu151 becomes protonated as the water molecule that helps to form the terminal peroxy- group on Zn_1 donates one of its hydrogens to it. The final two models that were studied both start with a neutral Glu151 while differing by a O^{2-} (Figure 4-6a) and a OH^- (Figure 4-6b) bridge.

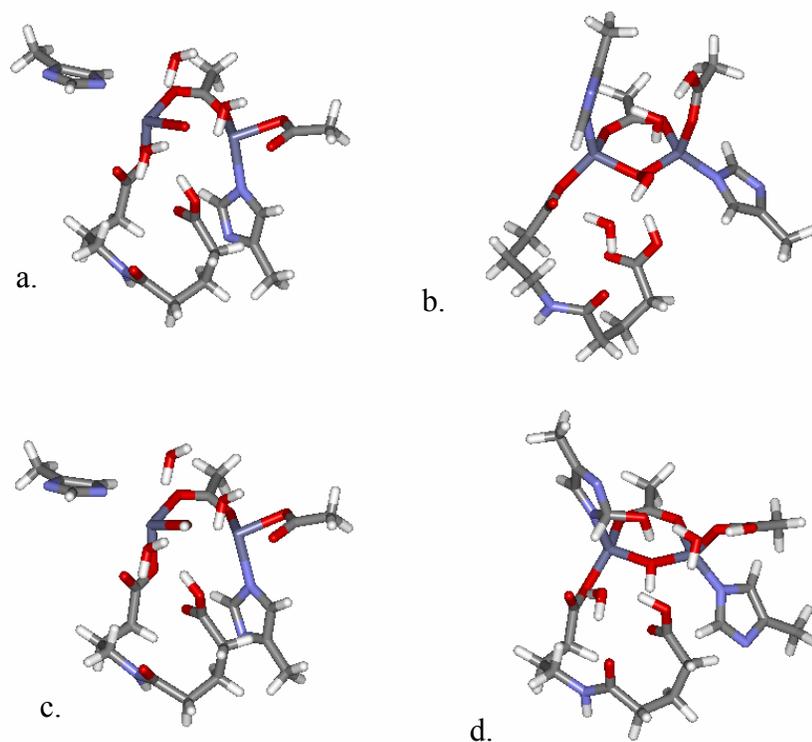


Figure 4-7. Initial structures (left) and B3LYP/6-31G* optimized geometries for models with Glu151 protonated. The starting structures vary only in the protonation state of the bridging group. Structure a) contains an O^{2-} bridge and the Zn ions in c) are bridged by a hydroxide group.

When model 4-6a is optimized, the formation of another peroxy- group is observed as a crystallographic water donates a hydrogen to Asp179 and the remaining OH^- binds to the original bridging O^{2-} . Glu151 remains protonated throughout the optimization. Structure 4-6b forms yet another interesting structure upon optimization. In this case, both Glu151 and the OH^- bridge retain their original protonation states. However, one of

the crystallographic water molecules donates a hydrogen to Asp179 while the remaining OH⁻ group complexes with His256.

Conclusions

Here, the initial efforts to detail the 1st-shell electronic structure, geometry, and protonation states of the active site of the aminopeptidase from *Aeromonas proteolytica* have been described. However, much work remains to be done until a complete picture of the mechanism of peptide hydrolysis in AAP can be revealed.

One purpose of this study was to investigate the different protonation states of the water/hydroxide bridge. To that end, many model active sites were created, each containing one of the three possible bridging species. In each case, the bridge interacted to some extent with the surrounding crystallographic waters in the active site. In some cases, a hydrogen-bonding network was established which helped to stabilize the active site structure. Some minimum energy structures also contained a peroxo- species that resulted from an oxo-bridge interacting with an active site water molecule. Finally, the DFT minimization calculations suggested that a hydroxide bridge was the most energetically stable, supporting some mechanistic studies previously done by other groups.

Another facet of the AAP study was to measure C-O bonds in the Zn-coordinating carboxylate residues. Each carboxylate side chain that complexes a Zn(II) ion can do so in either a monodentate or bidentate manner. Equivalent C-O bond lengths suggest that the electronic character of the side chain is distributed evenly throughout the carboxylate region and that each partially negative oxygen is interacting with a Zn cation. Residues in which one C-O bond is noticeably longer than the other indicate a residue that binds a metal cation in a monodentate fashion.

Finally, to gauge the importance of Glu151 as a proton acceptor in the initial stages of AAP peptide hydrolysis active sites were created with and without this residue. It was shown through DFT minimization that in species containing a OH⁻ or H₂O bridge between the Zn ions, that proton transfer occurred between the bridge and the previously unprotonated Glu151 residue. This further supports the notion that the bridging species must either be water or a hydroxide ion. Moreover, it suggests the necessity for an unprotonated Glu151 before substrate binding can occur.

This work has only scratched the surface of the computational work that can be performed on the AAP system. Other studies are currently underway to investigate the effects of 2nd-shell mutations around the active site. That study hopes to identify other key residues in active site geometry that may also participate in substrate binding or that may be targets for drug interactions. Some structures have been resolved which contain some small molecule substrate. Investigation of these structures could be used to better determine the electronic structure of substrate binding and locate any substrate interaction with 2nd-shell residues. Another study that is currently being performed is the full QM minimization of the implicitly solvated protein. In this work, the native unbound protein is being minimized along with structures containing 2nd-shell mutations. Comparing the minimized structures of the native protein and mutant proteins will illuminate the effects of the mutations on the overall structure of the active site.

CHAPTER 5 SURVEY OF DENSITY FUNCTIONAL THEORY METHODS

Introduction

The availability of large-scale parallel high-performance computer clusters is facilitating the application of *ab-initio* methods to large chemical systems such as biomolecules. Density Functional Theory (DFT) methods are a sensible choice for use in such calculations due to their relatively low expense compared to Hartree-Fock (HF) and post-HF methods and for the array of specific functionals which can be employed. However, when presented with a list of all of the DFT methods available, a scientist may only see an alphabet soup. Choosing an appropriate functional and basis set can be a daunting task, even for a seasoned computational chemist. One purpose of this survey is to evaluate a host of widely used DFT methods so that members of the scientific community at large can find which method is best suited to their needs.

Another way that the data presented in this study may be used is to quantify any progress made by recent DFT methods. The number of methods available to computational chemists has greatly increased over the last ten years and it seems that new functionals are being introduced every month in the literature. This work allows for the comparison of old “tried-and-true” methods to some of the newer functionals over a broad sampling of molecular properties.

Ultimately, this survey, which comprised more than 150,000 individual computational jobs, is the largest of its kind ever performed. Although some DFT methods have been omitted, a fair sampling of five families of DFT functionals is

presented and evaluated. The end result is a useful reference guide for future research on large scale organic and biomolecular systems using *ab initio* methods.

The next section in this chapter outlines the theory and development of *ab initio* techniques in a general sense from wave function methods to the most recent density functionals. Specific methods for calculating individual molecular properties are also discussed in the next section. Other sections of this chapter provide in-depth analysis of each facet of this work, addressing each molecular property in turn. The final section recapitulates the entire study with general conclusions.

Methods

All of the calculations in this work were performed using Gaussian 03 Rev C.01¹ and version D.01 and the functionalities therein. More detailed property-specific calculations are described below in individual sections. While the main focus of this work is to evaluate DFT methods, Hartree-Fock and second-order perturbation (MP2) methods are included for further comparison. A brief theoretical introduction to density functional methods is given in this section along with some discussion on the basis sets used in this work.

Schrödinger's equation can yield the exact energy of a system if the complete wave function and Hamiltonian are employed. However, a complete wave function and Hamiltonian are far too computationally expensive to be tractable and are difficult to define for multi-electron systems. A series of approximations have been adopted to simplify the Hamiltonian, thereby limiting the number calculations that must be performed on a system. A complete Hamiltonian for a system of N electrons takes the form:

$$H = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{1}{2M_A} - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (5-1)$$

The first term is the classical kinetic energy operator for the electrons. The third term is the Coulombic term for electrostatic interactions between the electrons and the nuclei, and the fourth term represents the charge repulsion between electrons. The second and fifth terms deal with nuclear kinetic energy and charge interactions, respectively, and are reduced to constants by the Born-Oppenheimer approximation which treats nuclei as fixed point charges in a field of moving electrons. The reduced form of the Hamiltonian after the B-O approximation is known as the electronic Hamiltonian of electrons moving in a field of fixed nuclear point charges:

$$H = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (5-2)$$

Following this approximation, the total energy of the system is the sum of the electronic energy and the constant nuclear charge interaction energy, which is dependent on the orientation of the nuclei to each other in space. A nuclear Hamiltonian can be used to account for motion of the nuclei as well. This simply consists of the second term from equation 5-1 and an added potential for nuclear motion.

The B-O approximation and its associated Hamiltonian satisfactorily describe the spatial parameters of the electron field. But to fully characterize an electron, spin must be taken into account. The concept of spin is roughly derived from the Pauli Exclusion Principle to ensure that no two electrons on an atom exist with the same energy or quantum configuration. A common visualization is that of one spin-up and one spin-down electron occupying a full orbital. In other words, one spatial orbital gives rise to

two unique spin orbitals. Electronic spin satisfies the notion of antisymmetry, which prohibits the existence of two like electrons in the same orbital. A spin orbital, χ_i , represents a complete picture of an electron both spatially and in terms of its spin.²

A new approximation is made to deal with the fully represented electrons, allowing for correct placement of the electrons into orbitals in a manner that satisfies the antisymmetry rule. When an antisymmetric wave function (equation 5-3), comprised of the spin orbitals of a ground state N-electron system, is operated upon by a Hamiltonian the lowest possible energy is returned (equation 5-4).

$$|\Psi_0\rangle = |\chi_1\chi_2\cdots\chi_N\rangle \quad (5-3)$$

$$E_0 = \langle \Psi_0 | \mathbf{H} | \Psi_0 \rangle \quad (5-4)$$

When E_0 is minimized with respect to the spin orbitals of Ψ_0 , the Hartree-Fock equation can be used to determine the optimal spin orbitals for the system:

$$f(i)\chi(\mathbf{x}_i) = \varepsilon\chi(\mathbf{x}_i) \quad (5-5)$$

This is the central tenet of the Hartree-Fock approximation and is the common starting point for more accurate quantum chemical methods. The Fock operator, $f(i)$, is a one-electron operator and $v^{HF}(i)$ in equation 5-6 is the effective potential incident upon electron i due to the other electrons in the system. In this representation, the many-body problem of electron-electron interaction has been reduced to a one-electron problem as electron-electron repulsion has been treated in an average manner.²

$$f(i) = -\frac{1}{2}\nabla^2 - \sum_{A=1}^M \frac{Z_A}{r_{iA}} + v^{HF}(i) \quad (5-6)$$

The field experienced by electron i is related to the spin orbitals of all other electrons in the system. This is referred to as electron correlation. In order to account for this relationship, the HF equation is solved in an iterative manner known as the self-consistent-field (SCF) method. In the SCF method, the spin orbitals are first described by an initial guess, from which an initial average electronic field can be calculated. The spin orbitals are subsequently modified until slight changes in the spin orbitals no longer affect the average field. At this point the energy has converged to the HF minimum. Using this method, however, can produce an infinite number of solutions to the HF equation since the number of accessible orbitals is theoretically very large. Basis functions are imposed on the system to limit the number of molecular orbitals that can be accessed by the electrons in the system. Different types of basis functions are described below, though in principle larger basis functions allow for more potential HF solutions to be explored in the attempt to increase the accuracy of the method. HF is a variational method and will always produce a higher energy than the ground state energy.

An example of a complete wave function method developed after HF is second order Møller-Plesset perturbation theory (MP2). Perturbation theory defines the total Hamiltonian as a zeroth-order term based on a Hamiltonian with known eigenfunctions and eigenvalues plus a perturbation term. The perturbation term may contain many orders of increasingly complex contributions to the exact Hamiltonian.

$$\mathbf{H} = \mathbf{H}_0 + \lambda\gamma \quad (5-7)$$

Equation 5-7 shows the general manner in which perturbation theory derives the exact Hamiltonian using an ordering parameter, λ , with a small perturbation γ . MP2 is a second-order technique which augments the zeroth-order Hamiltonian with elements of

two higher order perturbation terms derived from the Taylor expansion of the exact Hamiltonian in terms of the zeroth-order Hamiltonian.² The first order correction to the energy is the average of the perturbation Hamiltonian over the unperturbed wave function.³ This is called the variational method (or MP1), and the first-order correction to the energy is equal to the HF energy since a first-order correction is the variational integral for the HF wave function. The second-order expansion makes MP2 non-variational. The second-order correction to energy accounts for double excitations in the unperturbed wave function.³ Since the energy correction terms are derived from matrix elements of high order expansions, computational expense greatly increases with increasing perturbation order especially in terms of the amount of physical memory needed to complete the calculation. Consequently, MP2 typically scales as $O(\mathbf{n}^5)$ compared to HF which scales as $O(\mathbf{n}^4)$ where \mathbf{n} represents the number of orbitals in the system. Furthermore, MP2 is more sensitive to the quality of the initial guess than HF, which can result in energy convergence problems.

Density functional theory differs from the classical wave function methods described above because it uses the electron density, $\rho(\mathbf{r})$ as the primary variable instead of using the complete wave function $\psi(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2, \dots, \mathbf{r}_n, s_n)$. Application of DFT limits the system to its ground state. In practice, DFT provides a description of the ground state that is both mathematically less complicated and less computationally intensive than wave function methods.⁴ One benefit of DFT is that it is an exact method. In other words, when the exact wave function of a system is known, the DFT Hamiltonian can be applied to generate the exact ground state energy. With non-density functional methods, the external potential is used to determine the properties of a system. Hohenberg and Kohn

were the first to show that the ground-state electron density determines the Hamiltonian along with the ground-state wave function by replacing the external potential with electron density in the Schrödinger equation.^{5,6} With the ground-state Hamiltonian and wave function known, all electronic properties of the system can be calculated from the ground-state electron density. The density functional approach simplified the problem of solving the many-body Schrödinger equation to the minimization of a density functional.⁷ While the density functional can be minimized using several techniques, the Kohn-Sham approach⁸ is the most widely accepted. The density functionals evaluated in this work are all characterized as Kohn-Sham DFT methods.

DFT is not one method. Rather it encompasses a variety of methods that can be divided into several groups. One of the more difficult aspects of developing a density functional method is to devise a good approximation of the unknown energy density functional of the system.⁷ Different approaches to the inception of such approximations have led to different DFT families. Table 5-1 lists the thirty-seven density functional methods (divided into families) and the two wave function methods tested in this survey. Table 5-2 lists the basis sets that were paired with each of the thirty-seven functionals and two wave function methods (when computationally feasible) for each of the nine properties calculated in this work.

At this point, a brief overview of some of the DFT methods used in this study will be presented. More in-depth reviews exist such as Scuseria and Staroverov's 2005 essay⁷ and are suggested material for any interested reader. The Levine text provides an introductory glance at DFT and has been used to prepare this section.³

Table 5-1. The thirty-seven density functional method and two wave function methods tested in this survey with appropriate references.

Method	Reference	Method	Reference
HF	9	Hybrid-GGA	
MP2	10	B1LYP	11-13
		B3LYP	12-15
LSDA		PBE1PBE	16
SVWNV	17,18	B3P86	12,19
SPL	17,20	B3PW91	12,20-22
cSVWNV(0.3)	17,18,23	B98	24
GGA		Meta-GGA	
BLYP	12,13	VSXC	25
BPW91	12,20,21	BB95	12,26
PBELYP	13,16	MPWB95	26,27
PBEP86	16,19	TPSS	28,29
PBEPW91	16,20,21	MPWKCIS	26,30-32
PBEPBE	16	PBEKCIS	16,30-32
PW91LYP	13,20,21	TPSSKCIS	28-32
PW91P86	19-21		
PW91PW91	20,21	Hybrid-meta-GGA	
MPWLYP	13,19,27	BB1K	12,26,33
MPWP86	19,27	B1B95	12,26
MPWPW91	20,21,27	TPSS1KCIS	28-32,34
MPWPBE	16,27	PBE1KCIS	16,30-32,35
G96LYP	13,36	MPW1KCIS	11,26,30-32,37
G96P86	19,36		
HCTH	38		

Most DFT energy functionals include a correlation functional and an exchange functional. For some families of DFT methods an “exact exchange” term, similar in form to HF exchange (but employing K-S orbitals in DFT), is included in an effort to approximate the exact density functional. In this chapter the term “exact-exchange” refers to DFT exact exchange (which is not exact), not HF exact exchange (which is exact). Some of the more recently developed methods contain additional terms dependent on the kinetic energy density.

DFT methods can be categorized into many families based on the terms included in their respective functional forms. Members of five of these families are evaluated in this work. The simplest form of the density functional method is local spin density approximation (LSDA). This class of functional depends solely on the electron density. Generalized gradient approximation (GGA) methods account for reduced gradient of the electron density along with the electron density itself. Meta-GGA methods add a kinetic energy term to the GGA. Hybrid functionals are so-called because they mix an “exact exchange” term into their functional forms. Hybrid-GGA includes an “exact exchange” term along with a GGA functional, and hybrid-meta-GGA does the same for meta-GGA’s. The hybrid-meta-GGA family is the most recent to be developed and is functionally the most complex, comprising a GGA exchange functional, a GGA correlation functional, an “exact exchange” term, and a kinetic energy term.

The key approximations for density functional theory begin with the B-O simplified electronic Hamiltonian:

$$H = -\frac{1}{2} \sum_{i=1}^n \nabla_i^2 + \sum_{i=1}^n v(\mathbf{r}_i) + \sum_j \sum_{i>j} \frac{1}{r_{ij}} \quad \text{where} \quad v(\mathbf{r}_i) = -\sum_{\alpha} \frac{Z_{\alpha}}{r_{i\alpha}} \quad (5-8)$$

$v(\mathbf{r}_i)$ is the potential energy of the interaction between the electrons and the fixed nuclei in the system, which depends solely on the coordinates x_i , y_i , and z_i of electron i .

From Hohenberg and Kohn, it is known that the exact electronic energy can be determined from the ground state electron density probability, $\rho_0(\mathbf{r})$. Moreover, that the ground state energy is a functional of the ground state density:³

$$E_0 = E_v [\rho(\mathbf{r})] \quad (5-9)$$

The total ground state electronic energy can also be written as a sum of its component energies: the kinetic energy, the nuclei-electron interaction potential, and the electron-electron interaction potential. Since Hohenberg and Kohn proved that the ground state density can be used to determine all molecular properties, each of the energy components can also be expressed as functionals of the ground state density.³

$$E_0 = E_\nu[\rho_0] = \bar{T}[\rho_0] + \bar{V}_{Ne}[\rho_0] + \bar{V}_{ee}[\rho_0] \quad (5-10)$$

E_ν is dependent upon $\nu(\mathbf{r})$. So there is an established relationship between $\rho_0(\mathbf{r})$ and $\nu(\mathbf{r})$ such that:

$$\bar{V}_{Ne} = \langle \psi_0 | \sum_{i=1}^n \nu(\mathbf{r}_i) | \psi_0 \rangle = \int \rho_0(\mathbf{r}) \nu(\mathbf{r}) d\mathbf{r} \quad (5-11)$$

Now, one of the three terms in equation 5-10 is known, while the other two are not and must be approximated. The remaining two terms, $\bar{T}[\rho_0]$ and $\bar{V}_{ee}[\rho_0]$, can be combined into one term, $F[\rho_0]$, which is independent of $\nu(\mathbf{r})$ and whose value is unknown. In order to approximate $F[\rho_0]$, an initial guess density is required. In 1965, Kohn and Sham derived a method for finding ρ_0 and using it to determine E_0 .⁸

Kohn and Sham proposed an imaginary reference system of noninteracting electrons experiencing an external potential field, $\nu_s(\mathbf{r}_i)$. The external potential is chosen in such a manner that the corresponding ground-state density $\rho_s(\mathbf{r})$ of the reference state equals the ground-state density of the real system $\rho_0(\mathbf{r})$. Now that the reference density is known, the reference potential is functionally determined although its actual value may not be quantified.³

The Kohn-Sham one-electron Hamiltonian for the reference system of noninteracting electrons can be expressed as:

$$\mathbf{H}_{KS} = \sum_{i=1}^n \left[-\frac{1}{2} \nabla_i^2 + v_s(\mathbf{r}_i) \right] \quad (5-12)$$

And the reference system can be related to the real system by equation 5-13 in which a scaling variable λ is applied to the electron-electron interaction potential term and where v_λ is the external potential which will make $\rho_s(\mathbf{r}) = \rho_0(\mathbf{r})$.³

$$\hat{\mathbf{H}}_\lambda = \hat{\mathbf{T}} + \sum_i v_\lambda(\mathbf{r}_i) + \lambda \hat{\mathbf{V}}_{ee} \quad (5-13)$$

Kohn and Sham now define a kinetic energy term $\Delta \bar{\mathbf{T}}[\rho]$:

$$\Delta \bar{\mathbf{T}}[\rho] = \bar{\mathbf{T}}[\rho] - \bar{\mathbf{T}}_s[\rho] \quad \text{where} \quad \bar{\mathbf{T}}_s[\rho] = -\frac{1}{2} \langle \psi_s | \sum_i \nabla_i^2 | \psi_s \rangle \quad (5-14)$$

And an electron repulsion term $\Delta \bar{\mathbf{V}}_{ee}[\rho]$:

$$\Delta \bar{\mathbf{V}}_{ee}[\rho] = \bar{\mathbf{V}}_{ee}[\rho] - \frac{1}{2} \iint \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 \quad (5-15)$$

Now r_{12} is the distance between two electrons and the second term in equation 5-15 is the classical electron repulsion term for a distribution of density ρ of electrons. Having applied the Kohn-Sham method to our system, equation 5-10 becomes:

$$E_v[\rho] = \int \rho(\mathbf{r})v(\mathbf{r})d\mathbf{r} + \bar{\mathbf{T}}_s + \frac{1}{2} \iint \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 + \Delta \bar{\mathbf{T}}[\rho] + \Delta \bar{\mathbf{V}}_{ee}[\rho] \quad (5-16)$$

The last two terms of 5-16 are unknown, but they only constitute a small correction to the total energy. Together, the two terms comprise the exchange-correlation energy functional $E_{xc}[\rho]$. The total ground state energy can now be expressed as the sum of three density-dependent terms whose values are known and a fourth correction term

$E_{xc}[\rho]$ that is unknown. In order for a DFT method to have the ability to accurately calculate molecular properties, it must contain a good approximation to $E_{xc}[\rho]$.³

Earlier, the concept of spin-orbitals was introduced as they relate to the wave function of a system, and a wave function can be defined as being the Slater determinant of the spin-orbitals. By that definition, the electron probability density can be related to the sum of the squares of the Slater determinants of the spin-orbitals θ_i^{KS} of the imaginary system as shown in equation 5-17.

$$\rho_0 = \rho_s = \sum_{i=1}^n |\theta_i^{KS}|^2 \quad (5-17)$$

The Hohenberg-Kohn variational method was previously used to minimize the energy functional $E_v[\rho]$ by varying ρ . After establishing the relationship between ρ and θ_i^{KS} , it makes sense that the energy functional can also be minimized by varying θ_i^{KS} . Once the KS spin-orbitals have been defined, a new energy functional can be defined in a similar fashion as the Hartree-Fock equation (see equation 5-5). The new equation contains a kinetic energy term, a nuclei-electron interaction potential term, an electronic repulsion term, and a new function $v_{xc}(1)$ called the exchange-correlation potential.³

$$\left[-\frac{1}{2} \nabla_1^2 - \sum_{\alpha} \frac{Z_{\alpha}}{r_{1\alpha}} + \int \frac{\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + v_{xc}(1) \right] \theta_i^{KS}(1) = \varepsilon_i^{KS} \theta_i^{KS}(1) \quad (5-18)$$

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \quad (5-19)$$

$v_{xc}(\mathbf{r})$ is the functional derivative of the exchange-correlation energy E_{xc} . Now, if $E_{xc}[\rho]$ is known, then $v_{xc}(\mathbf{r})$ is also known. Each of these terms ultimately depend only on \mathbf{r} , the positions of the electrons in space.³

Unfortunately, the correct $E_{XC}[\rho]$ is not known. It can only be approximated. The Kohn-Sham method is widely used to estimate ν_{XC} and E_{XC} . Finally, different approaches to finding E_{XC} give rise to the different families of density functional methods. The first step in simplifying the problem of generating a reliable form of E_{XC} is to split it into an exchange-energy functional and a correlation-energy functional in this manner:

$$E_{XC} = E_X + E_C \quad (5-20)$$

E_X is formed based on the KS spin-orbitals, and E_C is the difference between E_{XC} and E_X . The magnitude of E_X is much larger than that of the correlation functional, but a good E_C is essential for an accurate DFT method nonetheless.³

With a general foundation of the components of density functionals now established, a brief synopsis of the five DFT families sampled in this study is presented. As mentioned, LSDA is the least complex density functional method used in this work as the exchange correlation functional depends only on the electron density. Spin-density functional theory, as proposed by Parr and Yang⁴ allows for electrons with different spins to be assigned separate spin-orbitals. In this manner, the electron densities of one type of spin-orbital α and those of another type of spin-orbital β are dealt with separately. The LSDA exchange-correlation functional takes the form:

$$E_{XC} = E_{XC}[\rho^\alpha, \rho^\beta] \quad (5-21)$$

This method treats the atomic environment as a uniform electron gas, and works best in systems where the density changes slowly. With that being said, LSDA functionals perform surprisingly well for equilibrium molecular geometries and vibrational frequencies, considering the method's similarity to unrestricted Hartree-Fock

(which performs poorly at each) and the reality that ρ does not change slowly in molecules. However, for most functionals of this kind, it has been seen that the accuracy of LSDA methods deteriorates when calculating thermodynamic properties.³

Accounting for the rapidly changing electron density is the first correction to the LSDA method. Generalized-gradient approximation (GGA) methods attempt to do this by including gradients of the densities ρ^α and ρ^β into their exchange-correlation functionals. Members of this functional class of may contain empirical parameters, such as the B88 exchange, or they may be non-empirical like the PW86 exchange.³

Hybrid functionals incorporate a HF-inspired “exact exchange” term in their exchange-energy functionals. The KS spin-orbital “exact exchange” is mixed with the gradient-corrected exchange and correlation functionals. Equation 5-22 gives the form of the “exact exchange” energy for KS spin-orbitals.

$$E_x = -\frac{1}{4} \sum_{i=1}^n \sum_{j=1}^n \langle \theta_i^{KS}(1) \theta_j^{KS}(2) \left| \frac{1}{r_{12}} \right| \theta_j^{KS}(1) \theta_i^{KS}(2) \rangle \quad (5-22)$$

The amount of KS “exact exchange” varies by method, and scaling constants placed with each term in a complete exchange-correlation energy functional can be varied to produce the highest accuracy attainable for a particular method. Such scaling constants are often tweaked to conform to some empirical data.³

The meta-GGA class includes a correction to the kinetic energy term $\Delta\bar{T}[\rho]$ in the attempt to further address issues not fully characterized by other density functional methods. The hybrid-meta-GGA class of functional is the most all-encompassing variation of the density functional method. From the basic spin-density approximation of the exchange-correlation energy, the hybrid-meta-GGA functionals add corrections for

rapidly changing electron density. They also mix in of a fraction of KS spin-orbital-based “exact exchange”, and some degree of correction to the kinetic energy. Surprisingly, there are hybrid-meta-GGA methods that are not empirically fit (or where the empirical bias is kept to a minimum) such as TPSSh.

Table 5-2. Basis sets employed in the survey of DFT methods.

Pople-type basis sets	Dunning-type basis sets
3-21G*	cc-pVDZ
3-21+G*	cc-pVTZ
6-31G*	cc-pVQZ ^a
6-31+G*	aug-cc-pVDZ
6-31++G*	aug-cc-pVTZ
	aug-cc-pVQZ ^a

a. cc-pVQZ and aug-cc-pVQZ used only for geometrical properties.

The basis sets used in this study range in size from small (3-21G*) to very large (aug-cc-pVQZ), and have all been widely validated.³⁹⁻⁴³ The Pople-style split-valence bases are the most familiar in quantum chemistry. Five of these basis sets were evaluated in this study: 3-21G*, 3-21+G*, 6-31G*, 6-31+G*, and 6-31++G*. The numbers in each basis set denote the number of Gaussian functions used to describe each orbital. For example, when the 3-21+G* basis function is applied, the inner shells are described in a single basis comprising a fixed linear combination of three Gaussian type orbitals (GTOs). The valence shells are described by a combination of inner valence and outer valence orbitals. The inner valence group is described by two fixed-combination GTOs, while one GTO is allowed to vary in the linear combination of atomic orbitals to represent the outer valence group. The + refers to diffuse functions which have been added to the heavy atoms. A ++ sign means that diffuse functions have been added to both heavy atoms and hydrogens.⁴⁴

Polarized basis sets, identified by *, are larger than normal basis sets. A single star (*) represents the addition of six Cartesian Gaussians (d_{xx} , d_{xy} , d_{xz} , d_{yy} , d_{yz} , and d_{zz}) or five pure d-orbitals (d_z^2 , d_{xy} , d_{xz} , d_{yz} , and $d_{x^2-y^2}$) to first row atoms in the molecule. A second star (**) adds a 2p function onto hydrogen atoms. The necessity for polarized functions derives from the distortion of the electron density of atoms once they form bonds, participate in other local interactions within molecules, or are acted upon by an external field. Although the magnitudes of these distortions may be small, the effects they have on the electronic energy are not negligible. Polarized basis sets attempt to replicate these distortions to more accurately predict molecular energy. Diffuse functions are a necessity for anions and atoms with lone pairs since regular basis functions do not adequately account for densely populated atomic orbitals far from the nucleus as exist in anions.⁴⁴ In effect, diffuse functions allow orbitals to occupy more space on the outer edges of the atoms.⁴⁵

The Dunning-style functions, known as correlation-consistent basis sets, follow a different manner of formalism and nomenclature. These functions are inherently polarized and account for effects related to molecular properties such as near-degeneracy of orbitals, and space and spin polarization. Dynamic correlations, such as orbital changes due to the movement of electrons, are also taken into consideration. Table 5-3 lists the polarization functions within each of the Dunning-type functions used in this survey as reproduced from the Gaussian 03 User's Reference.⁴⁶

Table 5-3. Valence shell polarization functions incorporated into the correlation-consistent basis sets of Dunning.

Atoms	cc-pVDZ	cc-pVTZ	cc-pVQZ
H	2s, 1p	3s, 2p, 1s	4s, 3p, 2d, 1f
B-Ne	3s, 2p, 1d	4s, 3p, 2d, 1f	5s, 4p, 3d, 2f, 1g
Al-Ar	4s, 3p, 1d	5s, 4p, 2d, 1f	6s, 5p, 3d, 2f, 1g

By augmenting a typical sp basis with higher order correlation effects, high accuracy basis sets are obtained. The inclusion of correlation effects, dynamic ones in particular, significantly adds to the computational expense of these methods as compared to Pople-type functions. Adding to the expense are the high angular momentum orbitals included in the valence shell polarization listed in Table 5-3. The Dunning-type functions employed in this study include: cc-pVDZ, cc-pVTZ, cc-pVQZ, and the diffuse versions of each of these: aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ (the “aug” qualifier replaces the + and ++ notation used in Pople-style basis sets).

Double zeta (DZ) and triple zeta (TZ) sets were used throughout the survey, but quadruple zeta (QZ) functions were only used in conjunction with the geometry test set due to the high computational expense of those basis sets. DZ, TZ, and QZ refer to the number of sizes of contracted Gaussian functions used in linear combinations of atomic orbitals to describe all of the molecular orbitals. For instance, a triple zeta basis employs three different sizes of contracted functions for each atomic orbital.

Computational Methods

The ability to accurately predict the ground state geometry of a molecule is one of the most important (and easily observed) aspects of a good computational method. Bond angles and bond lengths best describe molecular geometry. The test set for the evaluation of geometry performance comprises forty-four molecules containing the atoms C, H, O, N, S, and P. Within the set of molecules are seventy-one bond lengths and thirty-four bond angles. The test set is listed in Table 5-4. Ground state vibrational frequencies are closely related to molecular structure and are included in this portion of the study. Thirty-five molecules supply the 145 vibrational frequencies measured in our test set, which is tabulated in Table 5-5.

While this is the first study to encompass so many different DFT functionals and basis sets with a large test set, there have been many previous evaluations of DFT performance on molecular geometry and vibrational frequencies.^{23,47-55} Work done by Johnson used a group of small molecules from the G2 set to compare the performances of Slater and Becke type functionals along with HF and higher order methods.⁴⁹ Raymond and Wheeler investigated the accuracy of different Dunning-style basis sets at predicting the geometries of a set of small inorganic molecules.⁵² Wang and Wilson compiled a group of seventeen small molecules and compared the accuracy of several hybrid-GGA methods at several Dunning-type bases⁵⁴ while Riley *et al.* assessed LSDA and hybrid-GGA methods paired with Pople-type basis sets.²³

The methods for gathering geometrical parameters are fairly straightforward. Each molecule was optimized at each basis set/functional combination from the same initial structures. The initial structures were generated using WebLab Viewer Pro⁵⁶ in Cartesian coordinates or with the Molden⁵⁷ z-matrix editor. In Gaussian 03, the optimization and frequency calculations were performed with the default numerical grid and default energy and geometry convergence criteria. Previous studies have used geometries obtained at high levels of theory for use in frequency calculations with lower level methods. This work differs from those studies in that frequency calculations were carried out on optimized structures obtained by the same methods. This approach is appropriate for this study since high-level computational values are not available for large biomolecular systems.

The first ionization potential is the amount of energy required to remove one electron from a bound state to infinite separation, or the energy required to generate a

cation from an uncharged system. The IP test set is shown in Table 5-6 and contains thirty-seven small molecules, radicals, and ions.

Electron affinity is quantified as the energy gained by a neutral system when an unbound electron is captured, creating an anion. Table 5-7 lists the twenty-five molecules and radicals comprising the EA test set.

Finally, heat of formation is the difference between the enthalpy of a molecule and the sum of the energies of its individual atomic pieces. This quantity is a general measure of molecular stability and is used in predicting the energy released by reactions and in the calculation of other thermodynamic properties. A negative heat of formation typically indicates a stable molecule whose formation is spontaneous. A positive value indicates that an energy penalty has been paid in order to form the molecule from its elemental pieces. The heat of formation test set is the largest set in this survey and contains 127 singlet species (Table 5-8) and twenty-nine radicals (Table 5-9).

The molecules in the three test sets above were mainly gathered from the Gaussian G2/97 test set^{58,59}, though a few non-G2 molecules have been added to increase the number of phosphorus-containing compounds. The IP and EA test sets are completely derived from the G2/97 set with the exception of PO₂. Of the 156 members of the HOF test set, three molecules: PO₂, PH, and CH₃PH₂ are not from the G2/97 set.

Each of these molecular properties has been previously investigated using DFT in similar surveys.⁵⁸⁻⁶⁸ Curtiss *et al.* computed EA values for a set of fifty-eight molecules and IP for eighty-eight molecules using one LSDA, two GGA, and three hybrid-GGA DFT methods and the Gaussian-2 method with the 6-311+G(3df,2p) basis.⁵⁹ Ernzerhof and Scuseria completed a larger survey of the performance of SVWN, SVWNV, BLYP,

B3LYP, VSXC, PBEPBE, and PBE1PBE functionals with the 6-311+G(3df,2p) basis for atomization energies, IP, EA, and bond length.⁶⁴

Table 5-4. The bond lengths and bond angles test set.

#	Formula	Name	Bond	Angle
1	H ₂	hydrogen dimer	r(H-H)	
2 ^r	CH	Methyne	r(C-H)	
3	CH ₂ (¹ A ₁)	methylene singlet	r(C-H)	a(HCH)
4 ^r	CH ₂ (³ B ₁)	methylene triplet	r(C-H)	a(HCH)
5 ^r	CH ₃	Methyl	r(C-H)	a(HCH)
6	CH ₄	methane	r(C-H)	a(HCH)
7 ^r	NH	imidogen	r(N-H)	
8 ^r	NH ₂	amino	r(N-H)	a(HNH)
9	NH ₃	ammonia	r(N-H)	a(HNH)
10 ^r	OH	hydroxyl	r(O-H)	
11	OH ₂	water	r(O-H)	a(HOH)
12	HCCH	acetylene	r(C=C) r(C-H)	
13	H ₂ CCH ₂	ethene	r(C=C) r(C-H)	a(HCC) a(HCH)
14	H ₃ CCH ₃	ethane	r(C-C) r(C-H)	a(HCC) a(HCH)
15 ^r	CN	cyano	r(C=N)	
16	HCN	hydrogen cyanide	r(C=N) r(C-H)	
17	CO	carbon monoxide	r(C=O)	
18 ^r	HCO	formyl	r(C=O) r(C-H)	a(HCO)
19	H ₂ CO	formaldehyde	r(C=O) r(C-H)	a(HCH) a(HCO)
20	H ₃ COH	methanol	r(C-O) r(C-H _b) r(C-H _a) r(O-H)	a(OCH _a) a(COH) a(H _b CH _b) a(H _b CO)
21	N ₂	nitrogen dimer	r(N=N)	
22	H ₂ NNH ₂	hydrazine	r(N-N) r(N-H _a) r(N-H _b)	a(NNH _b) a(NNH _a) a(H _a NH _b)
23 ^r	NO	nitric oxide	r(N=O)	
24 ^r	O ₂	oxygen dimer	r(O=O)	
25	HOOH	hydrogen peroxide	r(O=O) r(O-H)	a(OOH)

r denotes non-singlet species

Table 5-4 (Continued): The bond lengths and bond angles test set.

#	Formula	Name	Bond	Angle
26	CO ₂	carbon dioxide	r(C=O)	
27	C ₃ H ₆	propene	r(C=C)	a(CCC)
			r(C-C)	a(HC ₃ C ₂)
			r(C3-H)	a(HC ₂ C ₃)
			r(C2-H)	a(HC ₂ C ₁)
			r(C1-H)	a(HC ₁ C ₂)
28	C ₃ H ₈	propane	r(C-C)	a(CCC)
			r(C2-H)	a(HC ₂ C ₁)
			r(C1-H)	a(HC ₁ C ₂)
29	C ₆ H ₆	benzene	r(C-C)	a(CCC)
			r(C-H)	a(HCC)
30	N(CH ₃) ₃	trimethyl amine	r(C-N)	a(CNC)
				a(HCN)
31	NH ₂ CH ₃	methyl amine	r(C-N)	a(HNC)
			r(N-H)	a(HNH)
32	CH ₃ OCH ₃	dimethyl ether	r(C-O)	a(COC)
				a(HCO)
33	NH ₂ CHO	formamide	r(C-N)	a(OCN)
			r(C-H)	a(OCH)
			r(C=O)	a(HNC)
			r(N-H)	
34	N ₂ O	nitrous oxide	r(N=N)	
			r(N=O)	
35 ^r	NO ₂	nitrogen dioxide	r(N=O)	a(ONO)
36	CH ₃ SH	thiomethanol	r(C-S)	a(HSC)
			r(S-H)	
37	CS ₂	carbon disulfide	r(C=S)	
38	SO ₂	sulfur dioxide	r(S=O)	a(OSO)
39 ^r	SN	sulfur nitride	r(S=N)	
40 ^r	S ₂	sulfur dimer	r(S=S)	
41 ^r	PH	phosphinidene	r(P-H)	
42 ^r	PH ₂	phosphino	r(P-H)	a(HPH)
43	PH ₃	phosphine	r(P-H)	a(HPH)
44	CH ₃ PH ₂	methyl phosphine	r(C-P)	a(HCP)
			r(C-H)	a(HPC)
			r(P-H)	

r denotes non-singlet species

Table 5-5. The test set for ground state vibrational frequency.

#	Formula	Name	Mode	Occurrence
1	H ₂	hydrogen molecule	Σ_g	1
2 ^r	O ₂	oxygen molecule	Σ_g	1
3	N ₂	nitrogen molecule	Σ_g	1
4 ^r	S ₂	sulfur molecule	Σ_g	1
5 ^r	CH	methylidyne	Σ	1
6	CH ₂ (¹ A ₁)	methylene singlet	A ₁	2
			B ₂	1
7 ^r	CH ₂ (³ B ₁)	methylene triplet	A ₁	1
			B ₂	1
8 ^r	CH ₃	methyl radical	A'	1
			A''	1
			E'	2
9	CH ₄	methane	A ₁	1
			E	1
			T ₂	2
10	C ₂ H ₂	ethyne	Σ_g	2
			Σ_u	1
			Π_g	1
			Π_u	1
11	C ₂ H ₄	ethene	A _g	3
			A _u	1
			B _{3g}	2
			B _{3u}	1
			B _{2g}	1
			B _{2u}	2
			B _{1u}	2
12	C ₂ H ₆	ethane	A _{1g}	3
			A _{1u}	1
			A _{2u}	2
			E _g	3
			E _u	3
13 ^r	NH	imidogen	Σ	1
14 ^r	NH ₂	amino radical	A ₁	2
			B ₂	1
15	NH ₃	ammonia	A ₁	2
			E	2

r denotes non-singlet species

Table 5-5 (Continued). The test set for ground state vibrational frequency.

#	Formula	Name	Mode	Occurrence
16	H ₂ N-NH ₂	hydrazine	A	7
			B	5
17 ^r	OH	hydroxyl radical	Σ	1
18	H ₂ O	water	A ₁	2
			B ₂	1
19	HO-OH	hydrogen peroxide	A	4
			B	2
20 ^r	PH	phosphinidene	Σ _g	1
21 ^r	PH ₂	phosphino radical	A ₁	2
			A ₁	2
22	PH ₃	phosphine	E	2
23 ^r	CN	cyano radical	Σ	1
24	HCN	hydrogen cyanide	Σ	2
			Π	1
25	CO	carbon monoxide	Σ	1
26	CO ₂	carbon dioxide	Σ _g	1
			Σ _u	1
			Π _u	1
			Π _u	1
27 ^r	HCO	formyl radical	A'	3
28	H ₂ C=O	formaldehyde	A ₁	3
			B ₂	2
			B ₁	1
29	CH ₃ OH	methanol	A'	8
			A''	4
30 ^r	NO	nitric oxide	Σ	1
31	CH ₃ SH	methanethiol	A'	8
			A''	3
32	CS ₂	carbon disulfide	Σ _g	1
			Σ _u	1
			Π _u	1
33	SO ₂	sulfur dioxide	A ₁	2
			B ₂	1
34 ^r	SN	nitrogen sulfide	Σ	1
35	CH ₃ PH ₂	methyl phosphine	A	13

r denotes non-singlet species

Table 5-6. The ionization potential test set.

#	Formula	Name
1	CH ₄	methane
2	NH ₃	ammonia
3	OH	hydroxy radical
4	H ₂ O	water
5	SH	mercapto radical
6	SH ₂ (² B ₁)	hydrogen sulfide (² B ₁)
7	SH ₂ (² A ₁)	hydrogen sulfide (² A ₁)
8	C ₂ H ₂	ethyne
9	C ₂ H ₄	ethene
10	CO	carbon monoxide
11	N ₂ (Σ _g)	nitrogen dimer Σ _g
12	N ₂ (Π _u)	nitrogen dimer Π _u
13	O ₂	oxygen dimer
14	S ₂	sulfur dimer
15	CO ₂	carbon dioxide
16	COS	carbonyl sulfide
17	CS ₂	carbon disulfide
18	CH ₂	methylene
19	CH ₃	methyl radical
20	CN	cyano radical
21	CH ₂ OH	hydroxymethyl radical
22	CH ₃ OH	methanol
23	CH ₃ SH	methanthiol
24	NH	imidogen
25	NH ₂	amino radical
26	C	carbon atom
27	N	nitrogen atom
28	O	oxygen atom
29	S	sulfur atom
30	H	hydrogen atom
31	CH ₃ PH ₂	methyl phosphine
32	P	phosphorous atom
33	PH	phosphinidene
34	PH ₂	phosphino radical
35	PH ₃	phosphine
36	P ₂	phosphorous dimer
37	PO ₂	phosphorous dioxide

Table 5-7. The electron affinity test set.

#	Formula	Name
1	C	carbon atom
2	O	oxygen atom
3	S	sulfur atom
4	CH	methyne
5	CH ₂ (³ B ₁)	methylene triplet
6	CH ₃	methyl radical
7	NH	imidogen
8	NH ₂	amino radical
9	OH	hydroxyl radical
10	SH	mercapto radical
11	O ₂	oxygen dimer
12	NO	nitric oxide
13	CN	cyano radical
14	S ₂	sulfur dimer
15	NCO	isocyanato radical
16	NO ₂	nitrogen dioxide
17	SO ₂	sulfur dioxide
18	CH ₃ O	methoxy radical
19	CH ₃ S	methylthio radical
20	CH ₂ CN	cyanomethyl radical
21	P	phosphorous atom
22	PO	phosphorous monoxide
23	PO ₂	phosphorous dioxide
24	PH	phosphinidene
25	PH ₂	phosphino radical

HOF, IP, EA and proton affinity were again studied by Curtiss et al. using the SVWN, BLYP, and B3LYP functionals and the Gaussian-3 method with the 6-311+G(3df,2p) basis function.⁶¹ The research performed by Brothers and Merz focused on the use of small basis sets (3-21G*, 3-21+G*, and MIDI!) with a variety of LSDA, GGA, and hybrid-GGA DFT methods to calculate heats of formation.⁶⁰ Brothers and Scuseria have recently presented a small basis set survey of heat of formation using the SVWN, cSVWNV, PBE, TPSS, and TPSSh functionals.⁶⁸

Table 5-8. The heat of formation test set, singlets only.

#	Formula	Name
1	CH ₂ (1A1)	methylene
2	CH ₄	methane
3	NH ₃	ammonia
4	OH ₂	water
5	SH ₂	dihydrogen sulfide
6	C ₂ H ₂	acetylene
7	C ₂ H ₄	ethene
8	C ₂ H ₆	ethane
9	HCN	hydrogen cyanide
10	CO	carbon monoxide
11	H ₂ CO	formaldehyde
12	H ₃ COH	methanol
13	N ₂	nitrogen dimer
14	H ₂ NNH ₂	hydrazine
15	HOOH	hydrogen peroxide
16	CO ₂	carbon dioxide
17	SC	carbon monosulfide
18	H ₃ CSH	methyl thiol
19	SO ₂	sulfur dioxide
20	COS	carbonyl sulfide
21	CS ₂	carbon disulfide
22	N ₂ O	nitrous oxide
23	O ₃	ozone
24	CH ₃ CCH	propyne
25	CH ₂ =C=CH ₂	allene
26	C ₃ H ₄	cyclopropene
27	CH ₃ CH=CH ₂	propylene
28	C ₃ H ₆	cyclopropane
29	C ₃ H ₈	propane
30	CH ₂ CHCHCH ₂	butadiene
31	C ₄ H ₆	2-butyne
32	C ₄ H ₆	methylene cyclopropane
33	C ₄ H ₆	bicyclobutane
34	C ₄ H ₆	cyclobutene
35	C ₄ H ₈	cyclobutane
36	C ₄ H ₈	isobutene
37	C ₄ H ₁₀	trans-butane
38	C ₄ H ₁₀	isobutane
39	C ₅ H ₈	spiropentane
40	C ₆ H ₆	benzene
41	CH ₃ NH ₂	methylamine
42	CH ₃ CN	methyl cyanide
43	CH ₃ NO ₂	nitromethane

Table 5-8 (Continued). HOF test set, singlets only

#	Formula	Name
44	CH ₃ ONO	methyl nitrate
45	HCOOH	formic acid
46	HCOOCH ₃	methyl formate
47	CH ₃ CONH ₂	acetamide
48	C ₂ H ₄ NH	aziridine
49	NCCN	cyanogen
50	(CH ₃) ₂ -NH	dimethyl amine
51	CH ₃ CH ₂ NH ₂	trans-ethylamine
52	CH ₂ CO	ketene
53	C ₂ H ₄ O	oxirane
54	CH ₃ CHO	acetaldehyde
55	HCOCOH	glyoxal
56	CH ₃ CH ₂ OH	ethanol
57	CH ₃ OCH ₃	dimethyl ether
58	C ₂ H ₄ S	thiirane
59	(CH ₃) ₂ -SO	dimethyl sulfoxide
60	C ₂ H ₅ SH	ethane thiol
61	CH ₃ SCH ₃	dimethyl sulfide
62	CH ₂ =CHCN	acrylonitrile
63	CH ₃ COCH ₃	acetone
64	CH ₃ COOH	acetic acid
65	(CH ₃) ₂ -CHOH	isopropanol
66	C ₂ H ₅ OCH ₃	methylethyl ether
67	N(CH ₃) ₃	trimethyl amine
68	C ₄ H ₄ O	furan
69	C ₄ H ₄ S	thiophene
70	C ₄ H ₅ N	pyrrole
71	C ₅ H ₅ N	pyridine
72	H ₂	hydrogen dimer
73	CH ₃ -CH=C=CH ₂	methyl allene
74	C ₅ H ₈	isoprene
75	C ₅ H ₁₀	cyclopentane
76	C ₅ H ₁₂	n-pentane
77	C ₅ H ₁₂	neopentane
78	C ₆ H ₈	1,3 cyclohexadiene
79	C ₆ H ₈	1,4 cyclohexadiene
80	C ₆ H ₁₂	cyclohexane
81	C ₆ H ₁₄	n-hexane
82	C ₆ H ₁₄	3-methyl pentane
83	C ₆ H ₅ CH ₃	toluene
84	C ₇ H ₁₆	n-heptane
85	C ₈ H ₈	1,3,5,7-cyclooctatetraene

Table 5-8 (Continued). HOF test set, singlets only

#	Formula	Name
86	C ₈ H ₁₈	n-octane
87	C ₁₀ H ₈	naphthalene
88	C ₁₀ H ₈	azulene
89	CH ₃ COOCH ₃	methyl acetate
90	(CH ₃) ₃ -COH	t-butanol
91	C ₆ H ₅ NH ₂	aniline
92	C ₆ H ₅ OH	phenol
93	C ₄ H ₆ O	divinyl ether
94	C ₄ H ₈ O	tetrahydrofuran
95	C ₅ H ₈ O	cyclopentanone
96	C ₆ H ₄ O ₂	benzoquinone
97	C ₄ H ₄ N ₂	pyrimidine
98	C ₂ H ₆ O ₂ S	dimethyl sulfone
99	N=C-CH ₂ -CH ₂ -C=N	1,2-dicyano ethane
100	C ₄ H ₄ N ₂	pyrazine
101	CH ₃ -C(=O)-C=CH	acetylene acetylene
102	CH ₃ -CH=CH-CHO	crotonaldehyde
103	CH ₃ -C(=O)-O-C(=O)-CH ₃	acetic anhydride
104	C ₄ H ₆ S	2,5-dihydrothiophene
105	(CH ₃) ₂ -CH-CN	isobutale nitrile
106	CH ₃ -CO-CH ₂ -CH ₃	methylethyl ketone
107	(CH ₃) ₂ -CH-CHO	isobutanal
108	C ₄ H ₈ O ₂	1,4-dioxane
109	C ₄ H ₈ S	tetrahydrothiophene
110	C ₄ H ₈ NH	tetrahydropyrrole
111	CH ₃ -CH ₂ -CH(CH ₃)-NO ₂	nitro-s-butane
112	CH ₃ -CH ₂ -O-CH ₂ -CH ₃	diethyl ether
113	CH ₃ -CH(OCH ₃) ₂	acetaldehyde dimethyl acetal
114	(CH ₃) ₃ C-SH	t-butanethiol
115	CH ₃ -CH ₂ -S-S-CH ₂ -CH ₃	diethyl disulfide
116	(CH ₃) ₃ C-NH ₂	t-butyl amine
117	C ₅ H ₆ S	2-methyl thiophene
118	C ₅ H ₇ N	n-methyl pyrrole
119	C ₅ H ₁₀ O	tetrahydropyran
120	CH ₃ -CH ₂ -CO-CH ₂ -CH ₃	diethyl ketone
121	C ₅ H ₁₀ O ₂	isopropyl acetate
122	C ₅ H ₁₀ S	tetrahydrothiopyran
123	cyc-C ₅ H ₁₀ NH	piperidine
124	(CH ₃) ₃ C-O-CH ₃	t-butyl methyl ether
125	(CH ₃) ₂ CH-O-CH(CH ₃) ₂	di-isopropyl ether
126	SO ₃	sulfur trioxide
127	CH ₃ PH ₂	methyl phosphine

Table 5-9. The heat of formation test set, radicals only.

#	Formula	Name
1	CH	methyne
2	CH ₂ (³ B ₁)	methylene
3	CH ₃	methyl
4	NH	imidogen
5	NH ₂	amino
6	OH	hydroxyl
7	CN	cyano
8	HCO	formyl
9	O ₂	oxygen dimer
10	S ₂	sulfur dimer
11	SO	sulfur monoxide
12	CCH	ethynyl
13	C ₂ H ₃ (2A')	vinyl
14	CH ₃ CO (2A')	acetyl
15	H ₂ COH (2A)	methoxy
16	CH ₃ CH ₂ O (2A'')	ethoxy
17	CH ₃ S (2A')	methylthio
18	C ₂ H ₅ (2A')	ethyl
19	(CH ₃) ₂ -CH (2A')	isopropyl
20	C(CH ₃) ₃	<i>t</i> -butyl
21	NO ₂	nitrogen dioxide
22	NO	nitric oxide
23	SH	mercapto
24	CH ₃ O	methoxy
25	C ₆ H ₅	phenyl
26	PH	phosphinidene
27	PH ₂	phosphino
28	PH ₃	phosphine
29	PO ₂	phosphorus dioxide

Literature values for IP, EA, and HOF were obtained from the G2/97 and G3 test set references for most molecules and from the NIST Chemistry WebBook for phosphorus-containing compounds. IP and EA values were calculated adiabatically using the methods described here. Ionization potentials were calculated by subtracting the total electronic energy of the initial system from that of the more positively charged final system. EA is derived in a similar fashion by subtracting the electronic energy of the initial molecule from that of the more negatively charged product. Heat of formation is

not explicitly listed in the Gaussian output, so it must be calculated using a combination of molecular and atomic energies.

In this study, HOF was calculated using the method described in the Gaussian white paper, “Thermochemistry in *Gaussian*”.⁶⁹ Heats of formation were calculated using the same basis sets that were used to optimize the structures. Due to high computational cost, optimization at the aug-cc-pVTZ level could not be completed. Instead, single point frequency calculations were performed using the aug-cc-pVTZ basis set on geometries obtained by the TPSS1KCIS/aug-cc-pVDZ method, which was found by a corresponding study to yield the most accurate geometries for this set of small molecules.⁷⁰

In order to calculate the heat of formation for a compound, the energies of the individual constituent elements must be calculated along with the thermochemical properties of the compound itself. The energies of the atoms are summed and subtracted from the thermally corrected enthalpy of the optimized molecule. For example, the heat of formation of methane would be:

$$\Delta H^f = \Delta H^{298}(\text{CH}_4) - \left(E_{\text{atom}}(\text{C}) + 4 \cdot E_{\text{atom}}(\text{H}) \right) \quad (5.23)$$

IP and EA calculations were performed geometry optimizations in Gaussian 03. For HOF calculations, the frequency calculations were also performed to generate enthalpy data. For most calculations, the default grid and convergence criteria were used. In some cases, a fine grid was specified. To resolve convergence problems, the quadratically convergent SCF method was implemented or the convergence criteria were relaxed slightly. The effects of implementing these convergence-aiding options were investigated, and the use of these non-default methods was found to have negligible effects on the properties being surveyed.

Hydrogen bonding is an important feature in many physical phenomena including the formation of clusters, protein folding and stability, and the formation of complexes between proteins and small molecules. The hydrogen bonding test set shown in Table 5-10 is made up of ten dimers involving H₂O, CH₃OH, H₂CO, NH₃, CH₃OCH₃, HCOOH, and HCONH₂. Experimental values for hydrogen bonding interaction energies are not readily available, so the reference values used for this test set were derived from CCSD(T) calculations performed by Tsuzuki and Lüthi.⁷¹

Table 5-10. The hydrogen bonding test set.

#	H-bonding system	E _{int} ⁷¹
1	H ₂ O - H ₂ O	-4.80
2	H ₂ O - MeOH	-4.90
3	H ₂ O - Me ₂ O	-5.51
4	MeOH - MeOH	-5.45
5	H ₂ O - H ₂ CO	-5.17
6	HCOOH - HCOOH	-13.93
7	NH ₃ - NH ₃	-2.94
8	NH ₃ - H ₂ O	-6.36
9	HCONH ₂ - H ₂ O	-8.88
10	HCONH ₂ - HCONH ₂	-13.55

E_{int} are given in kcal/mol.

A method's ability to accurately predict conformational energies is a measure of its worth as a technique for calculating geometric and electronic properties of molecules as well. The conformational energy test set is composed of ten small organic molecules each in two different conformations, and is listed in Table 5-11. Experimental values for conformational energies were obtained in studies performed by Zhao, Truhlar, Lynch, and González-García.^{34,65}

The study of reaction barrier heights is split into two segments. One part deals with small molecule reactions with radical (non-singlet) transition state species. The test set for small reaction barrier heights is listed in Table 5-12. The set consists of twenty-three

Table 5-11. The conformational energy test set.

#	Molecule	Conformations	$\Delta E^{34,65}$
1	Ammonia	planar vs. pyramidal	6.
2	ethylene	orthogonal vs. planar	65.
3	diethyl ether	tt vs tg	1.14
4	formamide	anti vs. eclipsed	20.
5	formic acid	Z vs. E	3.9
6	ethane	anti vs. eclipsed	2.9
7	methanol	anti vs. eclipsed	1.1
8	methylamine	anti vs. eclipsed	2.
9	methylcyclohexane	axial vs. equatorial	1.75
10	propene	anti vs. eclipsed	2.

ΔE are given in kcal/mol.

hydrogen-transfer reactions, whose experimental barrier heights are also listed in the table. The second group of barrier heights, listed in Table 5-13, contains larger reactions with singlet transition-state species. Most of these reactions were drawn from Jorgenson's PDDG paper.⁷²

Density functional methods have been used to study hydrogen-bonding interactions by several groups.^{35,71,73-80} Tsuzuki and Lüthi evaluated the BLYP, B3LYP, and PW91PW91 functionals, as well as the MP2 and HF methods, for the prediction of hydrogen bond interaction energies. Those studies were carried out using the Dunning type basis sets, cc-pVxZ (x=D, T, Q, 5). Zhao and Truhlar carried out studies to determine the accuracy of DFT methods for several types of non-bonding interactions: hydrogen bonding, charge transfer, dipole interaction, and the weak (dispersion) interaction. These studies were done using a very large number of functionals along with the 6-31+G(*d,p*), MG3S⁸¹ (a modified version of 6-31+G(3*d2f*, 2*df*, 2*p*), and aug-cc-pVTZ basis sets. To date, there have been only a limited number of studies concerned with the accuracy with which DFT methods predict conformational energies.^{81,82} Lynch, Zhao, and Truhlar evaluated the conformational energies of several conformer pairs of 1,2-

ethandiol and butadiene. These studies were done using a number of functionals based on the MPW correlation functional along with several basis sets. There have been a number of studies carried out to evaluate the accuracy with which DFT methods describe reaction barrier heights.^{34,83-86} In a recent study, Zhao, González-García, and Truhlar tested the accuracy of a large number of functionals along with several different basis sets for calculating the barrier heights of thirty-eight hydrogen transfer and thirty-eight non-hydrogen transfer reactions.³⁷

Table 5-12. The small system radical transition state reaction barrier test set.^{35,83-86}

#	Reaction	V _{forward}	V _{reverse}
1	OH + H ₂ ⇒ H + H ₂ O	5.7	21.8
2	CH ₃ + H ₂ ⇒ H + CH ₄	12.1	15.3
3	OH + CH ₄ ⇒ CH ₃ + H ₂ O	6.7	19.6
4	H + CH ₃ OH ⇒ CH ₂ OH + H ₂	7.3	13.3
5	H + H ₂ ⇒ H ₂ + H	9.6	n/a
6	OH + NH ₃ ⇒ H ₂ O + NH ₂	3.2	12.7
7	O + CH ₄ ⇒ OH + CH ₃	13.7	8.1
8	H + PH ₃ ⇒ PH ₂ + H ₂	3.1	23.2
9	H + OH ⇒ O + H ₂	10.7	13.1
10	H + H ₂ S ⇒ H ₂ + H ₂	3.5	17.3
11	NH ₂ + CH ₃ ⇒ CH ₄ + NH	8.0	22.4
12	NH ₂ + CH ₄ ⇒ CH ₃ + NH ₃	14.5	17.8

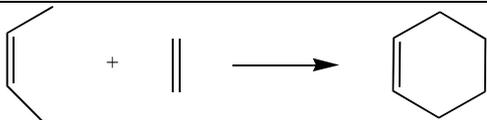
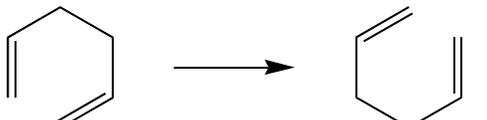
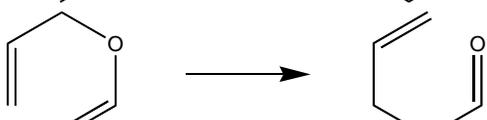
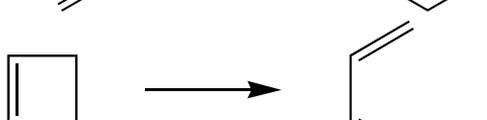
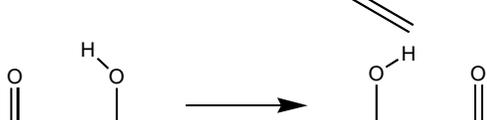
Barrier heights are listed in kcal/mol.

For the hydrogen bonding interaction energy calculations, the counterpoise method of Boys and Bernardi is employed in order to account for the basis set superposition error.⁸⁷ For each level of theory considered in this work the geometries of the hydrogen bond dimers are fully optimized on the counterpoise hypersurface. Then, the constituent monomers are also fully optimized using the same level of theory. Because of the difficulties associated with the extraction of the zero-point-exclusive binding energies from experimental data, binding energies obtained at a very high level of theory were

used as reference values. These reference values were determined at the CCSD(T) basis set limit by Tsuzuki and Lüthi.⁷¹

Experimental conformational energy reference data were all obtained from reference ⁷². These quantities are calculated as potential energy differences, that is, the difference in electronic energy between the most stable conformer and the least stable one. This is the method employed in several other studies.⁸⁸⁻⁹⁰

Table 5-13. The organic molecule singlet transition state reaction barrier test set.

Reaction	Exp.
	25.1 ⁷²
	33.3 ⁷²
	34.1 ⁷²
	32.9 ⁷²
	4-5 ⁷²
	35.4 ²³

Barrier heights are given in kcal/mol.

The barrier height reference data for the test set of small radical reactions are in the form of zero-point-exclusive, Born-Oppenheimer barrier heights. These barrier heights are simply calculated as the difference in electronic energy between the transition state

and the reactants. For the large singlet test set, the data are all directly from experiment so it is necessary to include vibrational effects into the calculation of the barrier heights. These barrier heights are calculated as the difference in the thermally corrected total enthalpy between the transition state and the ground state of the reactant(s). For all six reactions in the singlet set, initial coordinates of the transition states were constructed to have sensible transition state-like geometries. The transition state optimization method in Gaussian 03 was implemented. Reaction barrier heights were obtained from optimized reactant and transition state structures for all functional and basis set combinations except MP2/aug-cc-pVTZ.

Ideally, the theoretical determination of a particular property should be calculated at the level of theory at which the geometry of the system is obtained. Most of the calculations in this work have been completed in just such a fashion. However, because of substantial problems associated with transition state convergence in the investigations involving the small radical barrier height test set, the geometries determined by Lynch and Truhlar at the QCISD/MG3 level were used for all calculations of this property (see <http://comp.chem.umn.edu/truhlar/>).⁸⁶

Results by Property

Bond Lengths

The average unsigned bond length errors for the gradient corrected, GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals, with Pople type basis sets are given in Figure 5-1. HF, MP2, and LSDA results are given in Table 5-14. Inspection of these figures reveals that the hybrid-GGA and hybrid-meta-GGA methods generally yield the lowest bond length errors when paired with these basis sets. The GGA and meta-GGA functionals both yield similar results with errors that are slightly greater than those

of the hybrid and hybrid-meta functionals. Among the LSDA functionals, SVWN5 and SPL obtain errors that are comparable to those of GGA and meta-GGA while c-SVWN5 yields errors that are higher than the other LSDA methods. Of the GGA functionals, from the meta-GGA class, the VSXC functional yields the lowest unsigned bond length errors.

Figure 2 gives average unsigned bond length errors for the gradient corrected methods along with the Dunning type basis sets. It can be seen that, overall, the hybrid-GGA functionals yield the lowest errors. Meta-GGA functionals and hybrid-meta-GGA functionals perform nearly as well, followed by the GGA functionals. LSDA methods yield errors that are generally higher than those of GGA functionals. It should be noted that, for the smallest of the Dunning type basis sets, cc-pVDZ and aug-cc-pVDZ, the hybrid-GGA and hybrid-meta-GGA methods generally outperform all other methods by a significant margin. For the larger, cc-pVTZ, cc-pVQZ, aug-cc-pVTZ, and aug-cc-pVQZ, basis sets the hybrid-GGA, hybrid-meta-GGA, and meta-GGA functionals all yield, on average, comparable results. As in the case of the Pople type basis sets, VSXC yields the lowest bond length errors.

Generally, increasing the basis set size results in smaller bond length errors, the most dramatic improvements can be seen when the small 3-21G* and 3-21+G* basis sets are compared to the 6-31G* and 6-31+G* basis sets. The addition of diffuse functions to hydrogen atoms in the 6-31++G* makes little difference in the magnitudes of bond length errors. For the Dunning type bases, the addition of diffuse functions improves accuracy only slightly. It is interesting to note that for all of the functionals considered here, the bond length errors obtained using the 6-31G* and 6-31+G* basis sets are lower than

those obtained using the much more computationally expensive cc-pVDZ and aug-cc-pVDZ basis sets.

One of key feature of the bond length data is that, there is little variation within a particular functional class for a given basis set. For example, the average unsigned bond length errors calculated with the hybrid-GGA functionals along with the 6-31G* basis set are all within 0.002Å of one another. This trend is especially evident for the GGA, hybrid-GGA, and meta-GGA functionals whereas there is quite a bit more variation within the LSDA and hybrid-meta-GGA class of functionals. Another aspect of this trend is that, typically, the amount of variation within a particular functional class decreases as the basis set size increases. Individual functionals that deviate significantly from the other functionals within their class are c-SVWN, HCTH, B98, VSXC, and BB1K.

Table 5-14. Bond lengths, bond angles, and vibrational frequencies for HF, MP2, and LSDA methods.

Method	3-21G*	3-21 +G*	6- 31G*	6-31 +G*	6-31 ++G*	cc- pVDZ	cc- pVTZ	cc- pVQZ	aug-cc- pVDZ	aug-cc- pVTZ	aug-cc- pVQZ
Bond lengths (Å)											
HF	0.015	0.016	0.015	0.014	0.014	0.013	0.018	0.019	0.012	0.017	0.019
MP2	0.024	0.027	0.011	0.012	0.012	0.014	0.009		0.015	0.009	
SVWNV	0.024	0.024	0.018	0.017	0.017	0.021	0.014	0.014	0.019	0.014	0.014
SPL	0.025	0.025	0.018	0.017	0.017	0.022	0.014	0.014	0.019	0.014	0.014
cSVWNV	0.034	0.035	0.025	0.025	0.025	0.029	0.019	0.018	0.027	0.019	0.018
Bond Angles (degrees)											
HF	1.85	2.45	1.45	1.55	1.54	1.40	1.50	1.51	1.48	1.53	1.53
MP2	1.55	2.02	1.31	1.39	1.37	1.65	1.22		1.26	1.17	
SVWNV	1.80	2.58	1.36	1.27	1.27	1.60	1.25	1.22	1.28	1.24	1.23
SPL	1.80	2.55	1.37	1.27	1.27	1.60	1.25	1.22	1.27	1.23	1.22
cSVWNV	1.69	2.31	1.52	1.28	1.28	1.75	1.24	1.20	1.25	1.19	1.17
Vibrational frequencies (cm ⁻¹)											
HF	210	203	236	235	234	211	209	203	207		
MP2	151	147	149	141	141	126	122	113	117		
SVWNV	83	87	51	50	50	58	51	59	52		
SPL	82	87	52	50	50	57	52	58	51		
cSVWNV	93	100	58	57	58	75	63	71	65		

For singlets, the hybrid-GGA and hybrid-meta-GGA methods yield the lowest errors when paired with most basis sets. However, for the TZ bases hybrid-GGA, meta-GGA, and hybrid-meta-GGA methods yield the best singlet results. For the small Pople type basis sets, 3-21G* and 3-21+G* the LSDA methods yield results that are slightly better than those of the GGA functionals. At higher basis sets, the all other density functional classes outperform the LSDA methods. In terms of radical systems, the hybrid-GGA and hybrid-meta-GGA are the density functional methods that yield the lowest bond length errors for all basis sets, since the best results overall are obtained by MP2. For all basis sets, the meta-GGA methods obtain errors that are higher than those of the hybrid-GGA and hybrid-meta-GGA methods.

For all methods except MP2, single and triple bonds are calculated much more accurately than double bonds. For all basis sets the LSDA, GGA, and meta-GGA methods tend to yield lower unsigned errors for triple bonds than for double bonds. Hybrid-GGA and hybrid-meta-GGA also obtain lower errors for triple bonds (compared to single bonds) when paired with Pople type and cc-pVDZ/aug-cc-pVDZ basis sets, but give higher errors for triple bonds with the larger Dunning type basis sets. Hartree-Fock and MP2 generally obtain lower bond length errors for single bonds than triple bonds.

Bond Angles

The average unsigned bond angle errors are shown in Figures 5-3 and 5-4 and in Table 5-14. Generally, the hybrid-GGA and hybrid-meta-GGA methods produce the lowest errors. The best results among the small basis sets are generally obtained with the hybrid-GGA functionals along with the 3-21G* basis set. The hybrid-meta-GGA functionals also yield very good results although there is some of variation within this class. The LSDA methods produce bond angle errors that are higher than most of the

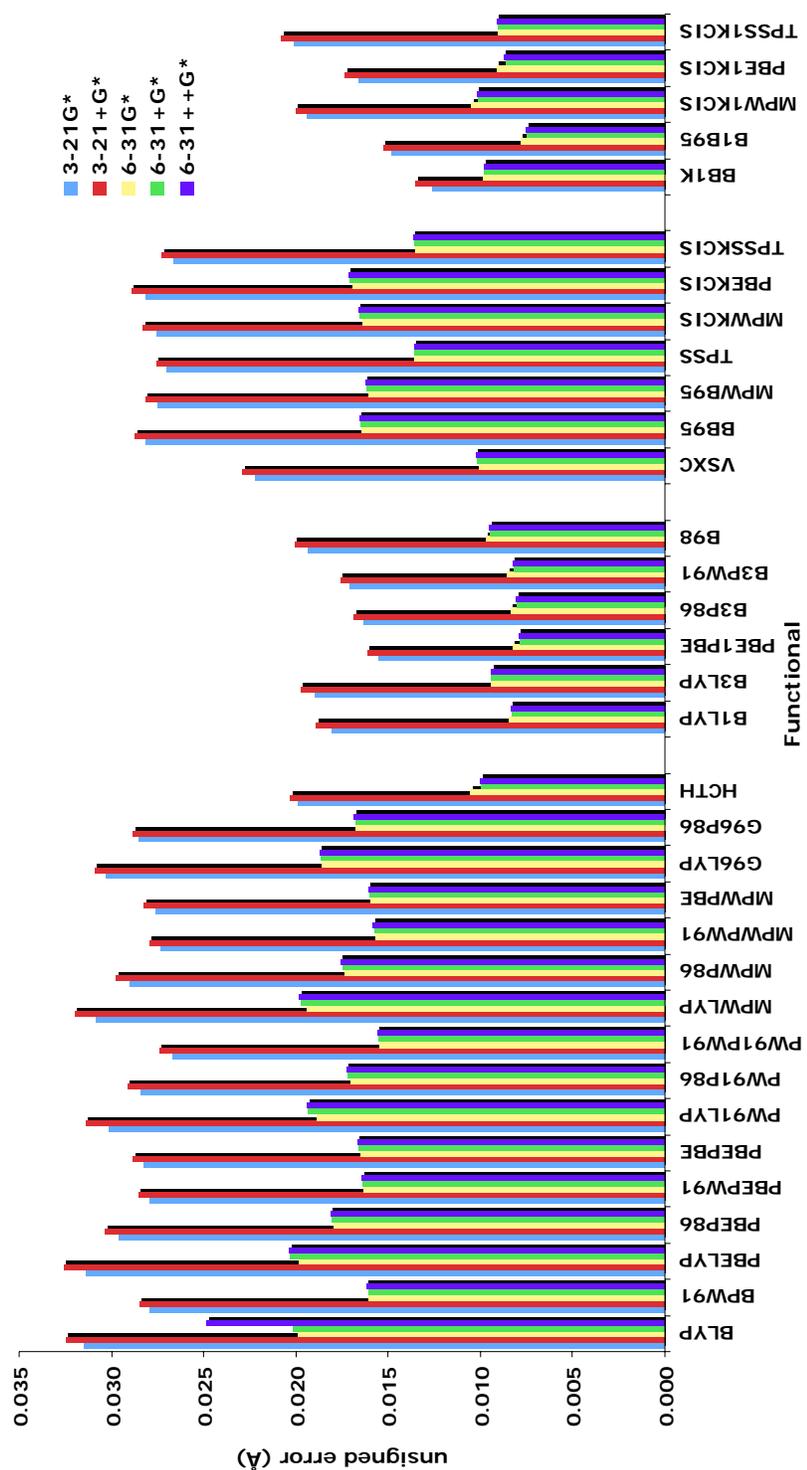


Figure 5-1. Average unsigned bond length errors (in Å) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Pople type basis sets.

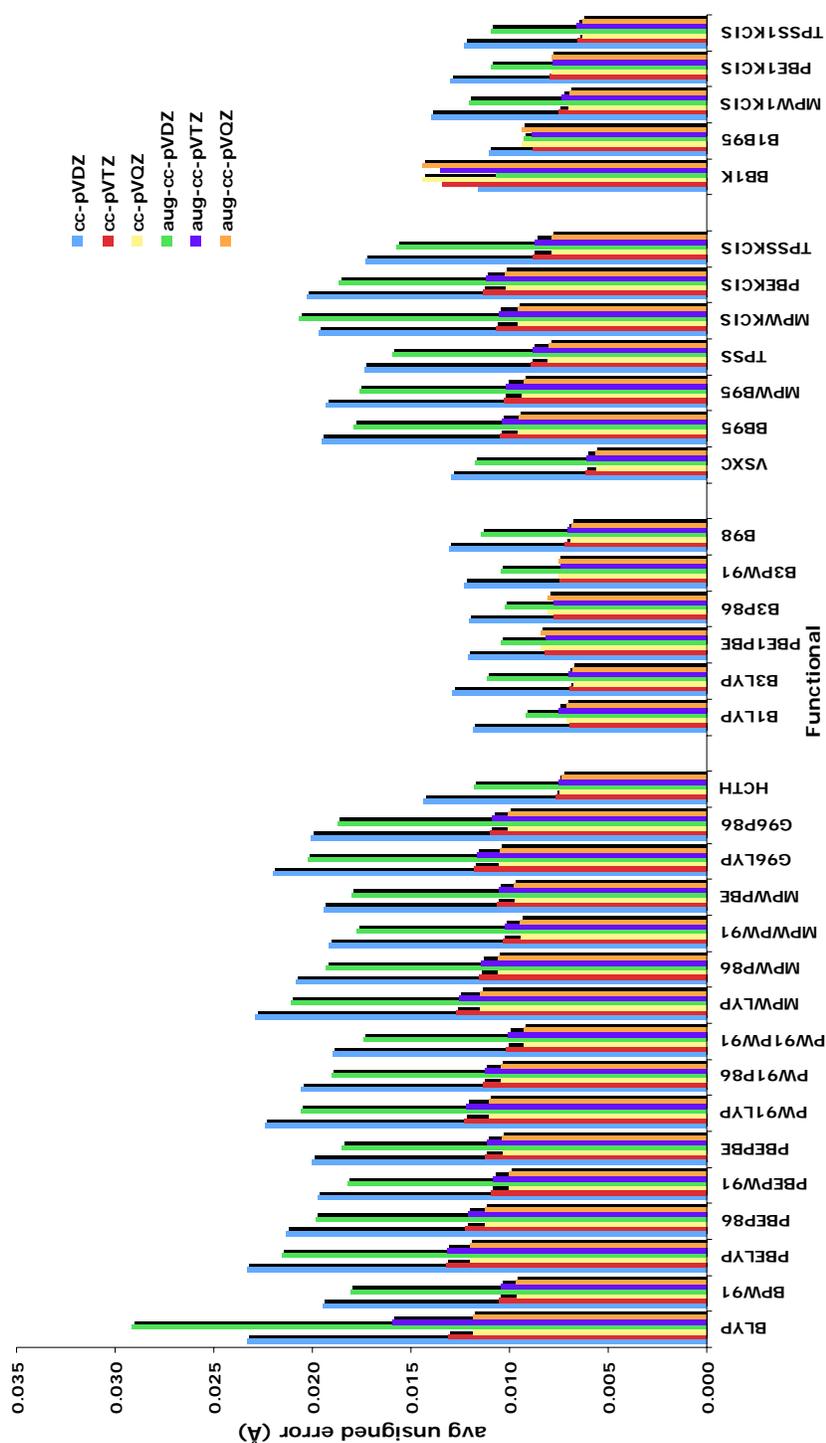


Figure 5-2. Average unsigned bond length errors (in Å) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Dunning type basis sets.

gradient-corrected functionals. The addition of diffuse functions to the basis set results in a significant increase in the average bond angle errors for these functionals. The SVWN5 and SPL functionals both yield very similar errors while the c-SVWN5 functional obtains results that are only slightly better than its LSDA counterparts when paired with the small bases. Overall, the best small basis set results are obtained by the B1LYP/3-21G* and PBE1PBE/3-21G* methods which both calculate the average unsigned bond angle error to be 1.36°.

The larger Pople basis sets, 6-31G*, 6-31+G*, and 6-31++G*, generally produce better bond angle results for the basis sets containing diffuse functions. The wave function based methods yield lower errors when paired with the 6-31G* basis set. It should also be noted that the addition of diffuse functions to hydrogen atoms in the 6-31++G* basis set does not result in any significant improvement over the 6-31+G* basis. Once again, the hybrid-GGA and hybrid-meta-GGA methods generally give the lowest unsigned bond angle errors. As in the case of the small Pople basis sets, there is a great deal of variation in the meta-GGA class of functionals. LSDA methods all yield similar results when paired with basis sets containing diffuse functions. However, c-SVWN5 produces errors that are significantly higher than those of SVWN5 and SPL when paired with the 6-31G* basis set. For 6-31G*, HF generally outperforms the GGA and meta-GGA methods and yields errors that are higher than those of all hybrid-GGA and hybrid-meta-GGA methods.

One aspect of the data for the Dunning type basis sets is that the errors obtained with the cc-pVDZ basis set are much higher than those of all other Dunning basis sets for all methods considered except Hartree-Fock. In fact, cc-pVDZ is generally outperformed

by all other basis sets, with the exception of 3-21+G*, for all functional methods. The hybrid-GGA and hybrid-meta-GGA functionals generally yield the lowest unsigned bond angle errors for the Dunning type basis sets. Generally speaking, increasing the basis set size results in lower unsigned errors, this trend is especially evident for the GGA and meta-GGA functionals paired with non-diffuse basis sets. When diffuse functions are added to the cc-pVTZ basis set there is a significant increase of accuracy for GGA and meta-GGA functionals. For most functionals there is a small decrease in bond angle error upon addition of diffuse functions to cc-pVQZ. Overall, the lowest error of 1.11° is obtained by the hybrid-GGA PBE1PBE/aug-cc-pVQZ method. PBEP86 and MPWP86 are the GGA functionals that typically yield the lowest unsigned bond angle errors, while HCTH is the least accurate in this class. It should be noted that BLYP exhibits trends that are quite different than those of other GGA functionals. BLYP/aug-cc-pVDZ yields errors that are much higher than all other GGA/aug-cc-pVDZ combinations while BLYP/aug-cc-pVTZ produces errors that are significantly lower than all other GGA/aug-cc-pVTZ combinations. LSDA methods all yield similar results with the exception of c-SVWN/cc-pVDZ. HF performs better when paired with the non-diffuse basis sets and larger Dunning-type bases generally yield larger errors than the smaller Dunning bases. The unsigned bond angle errors obtained with MP2 improve with increasing basis set size and with the addition of diffuse functions to the basis set.

HF, MP2, LSDA, GGA, and meta-GGA methods obtain lower bond angle errors for singlet states for most Dunning and large Pople type basis sets. By and large, the hybrid-GGA and hybrid-meta-GGA methods yield lower errors for radicals than for singlets. Most methods produce better results for radical species when paired with the 3-

21+G* basis set. For singlets, the LSDA functionals produce the lowest unsigned bond angle errors among all functional based methods when paired with the Dunning type basis sets except cc-pVDZ and with the 6-31+G* and 6-31++G* Pople type basis sets. Hybrid-GGA methods give the lowest errors among DFT methods for 3-21G* and 6-31G*. In terms of radical species, the hybrid-GGA methods give the lowest unsigned bond angle errors for all basis sets considered in this work. Hybrid-meta-GGA methods generally produce errors that are slightly larger than those of hybrid-GGA functionals.

Ground State Vibrational Frequencies

Figures 5-5 and 5-6 show the average unsigned vibrational frequency errors for gradient corrected DFT methods. Vibrational errors for the LSDA methods as well as HF and MP2 are given in Table 5-14. It has been observed previously that frequency errors for MP2 and HF methods are larger than those of most DFT methods. It is observed in this study that the hybrid-GGA and hybrid-meta-GGA functionals, which include a Hartree-Fock like exchange term, are less accurate than other classes of DFT functionals in predicting harmonic vibrational frequencies. As expected, HF and MP2 errors are far worse than errors obtained with DFT calculations. Compared with the GGA, meta-GGA, and LSDA classes of functionals, MP2 errors are higher by a factor of about two or three, while HF errors are typically about three to five times higher. The lowest vibrational frequency error obtained by the MP2 method is 113 cm^{-1} for MP2/aug-cc-pVDZ while its highest error is 151 cm^{-1} for MP2/3-21G*. HF errors for vibrational frequencies range from 203 cm^{-1} , for HF/3-21+G*, to 236 cm^{-1} , for HF/6-31G*

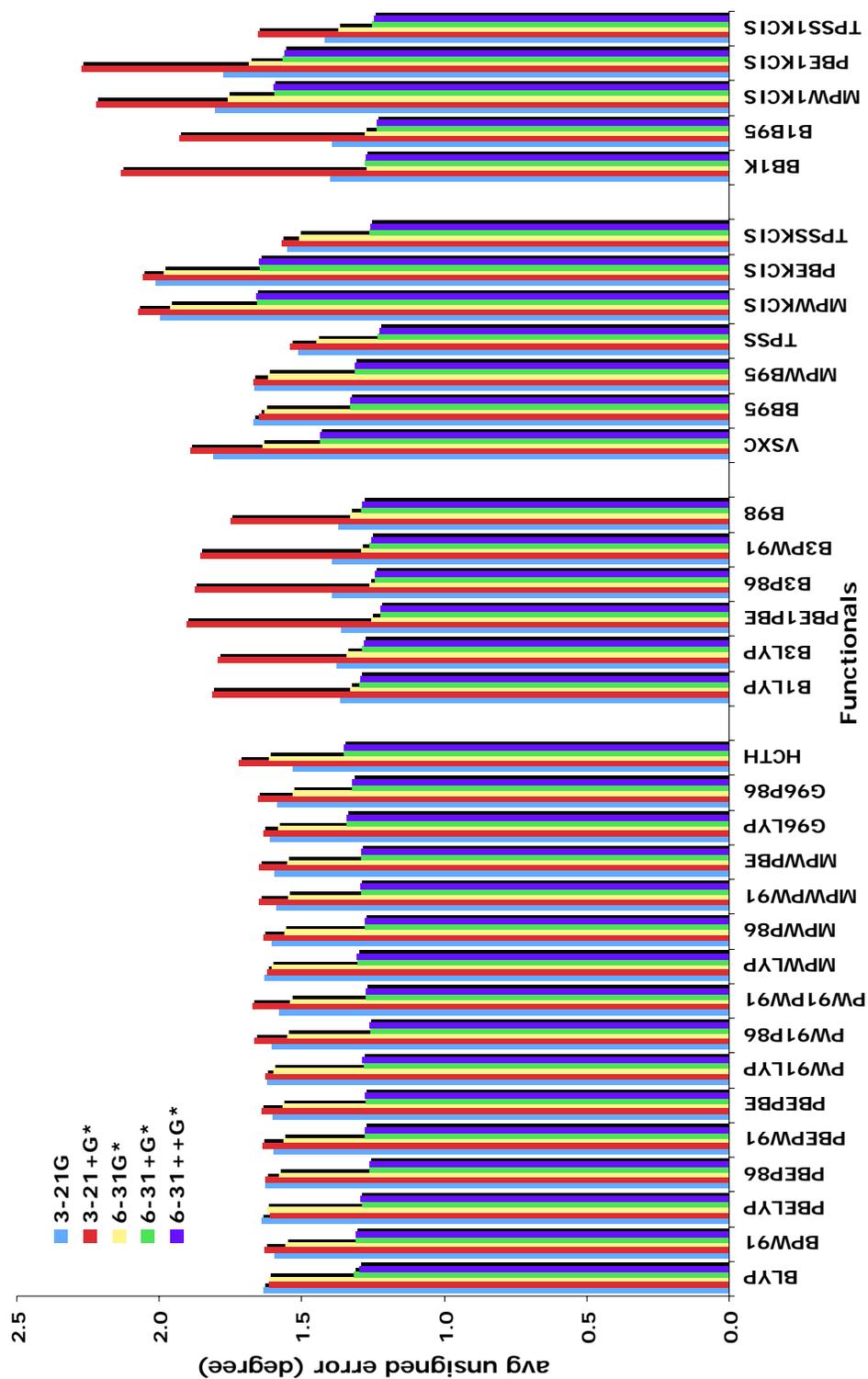


Figure 5-3. Average unsigned bond angle errors (in degrees) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Pople type basis sets.

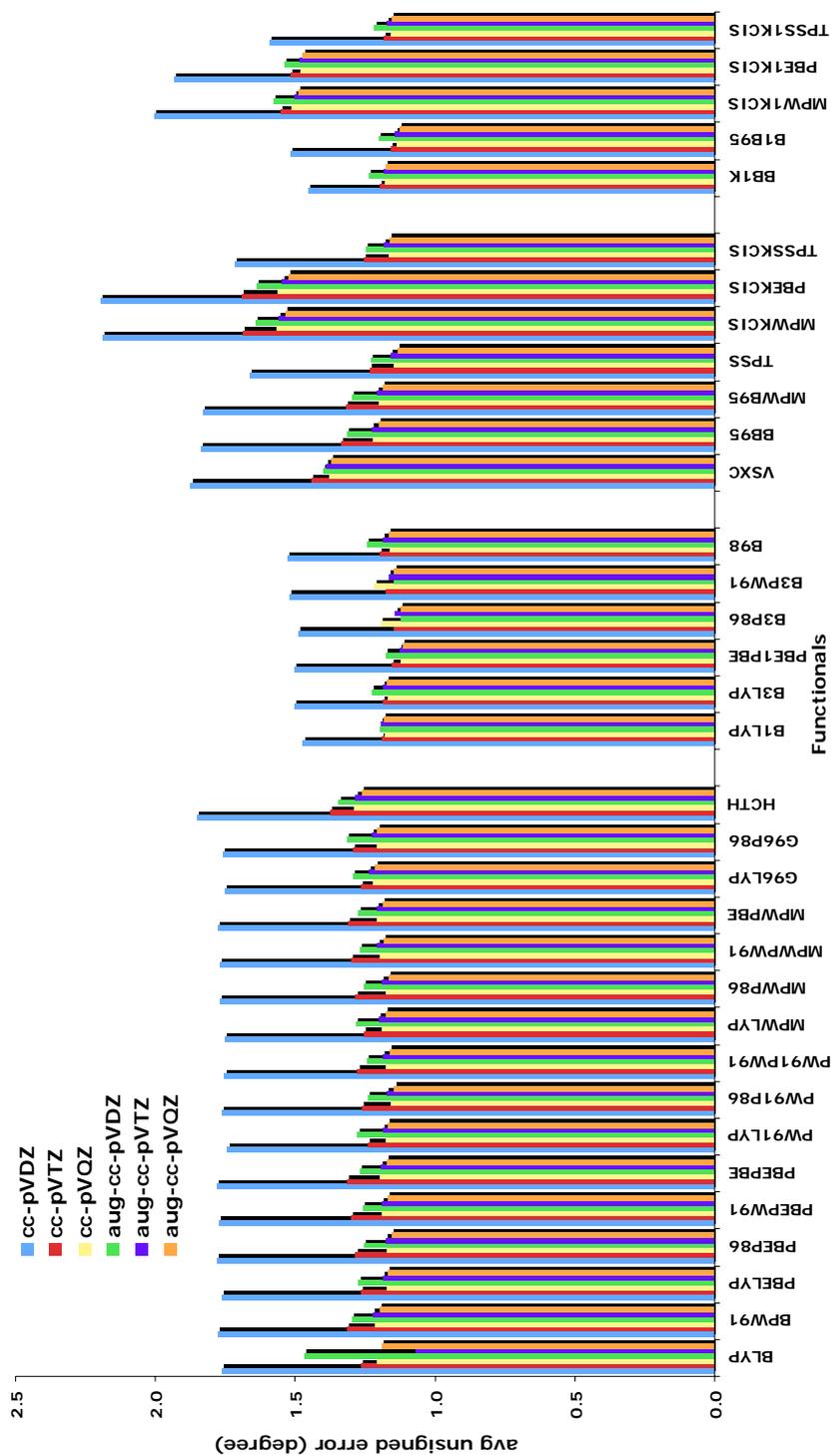


Figure 5-4. Average unsigned bond angle errors (in degrees) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Dunning type basis sets.

The inclusion of diffuse functions in the basis set does not greatly affect the ability of the functionals to predict vibrational frequencies. As shown in the figures, 3-21G* and 3-21+G* are nearly identical in performance. There is only a small improvement, on the order of 5 cm^{-1} , with the use of the 6-31+G* and 6-31++G* basis sets as opposed to 6-31G*. Augmented correlation-consistent basis sets do not perform markedly better than their non-diffuse counterparts for the double-zeta and triple-zeta basis sets.

The three LSDA functionals yield lower average errors than hybrid-meta-GGA and hybrid-GGA functionals, but they are generally higher than the errors of meta-GGA and GGA functionals. The SPL and SVWN5 functionals perform slightly better than c-SVWN5, with SPL performing the best of the three. The 6-31+G* and 6-31++G* basis sets yield the lowest errors for the SPL functional, around 50 cm^{-1} .

Within the GGA class, there is little variation of performance between functionals with the exception of HCTH, which yields vibrational frequency errors that are significantly higher than any of the other functionals. In this class the Dunning style triple-zeta basis sets give the lowest unsigned vibrational frequency errors. Both the diffuse and non-diffuse variants of this basis set produce unsigned errors of $40\text{-}44\text{ cm}^{-1}$. It is interesting to note that the 6-31+G* and 6-31++G* basis sets, which are much less computationally expensive than the cc-pVTZ and aug-cc-pVTZ basis sets, yield errors that are only $5\text{-}8\text{ cm}^{-1}$ higher than these Dunning type basis sets. 6-31+G* and 6-31++G* also give results that typically outperform the cc-pVDZ and aug-cc-pVTZ basis sets. Of the 16 functionals in the GGA family, we find that MPWLYP and MPWP86 perform the best, but their advantage over most of the other functionals is only slight.

B3LYP is the most accurate of the hybrid-GGA class for calculating vibrational frequencies. B3LYP/cc-pVDZ yields the lowest error in this class at 70 cm^{-1} . PBE1PBE does not perform as well as the other members of this class. The meta-GGA class also shows some variance between functionals. VSXC, which is among the best meta-GGA methods for calculating bond lengths, performs poorly for vibrational frequencies. BB95, MPWB95, MPWKCIS and PBEKCIS are similar in performance throughout. BB95/aug-cc-pVTZ and MPWB95/aug-cc-pVTZ yield the lowest error of the class at 43 cm^{-1} . Finally, the hybrid-meta-GGA group does not perform as well as most other classes due to its inclusion of the HF exchange. TPSS1KCIS and MPW1KCIS are the best functionals in this functional class, the cc-pVTZ and aug-cc-pVTZ basis sets typically give the lowest average errors within this group.

Ionization Potential

The average ionization potential unsigned errors for each functional/basis set combination are given in Figures 5-7 and 5-8 and in Table 5-15. Table 5-15 gives results of HF, MP2, LSDA functionals, and B3P86 (which yields high errors) for all basis sets. The best result for ionization potentials is obtained with the hybrid-meta-GGA functional B1B95 combined with the aug-cc-pVTZ basis set yielding an error of 4.25 kcal/mol. The worst average unsigned error among the density functional methods is 15.05 kcal/mol and is obtained using the hybrid-GGA functional B3P86 along with the 3-21+G* basis set. Overall, the largest unsigned error of 26.53 kcal/mol is obtained with the HF/3-21G* method.

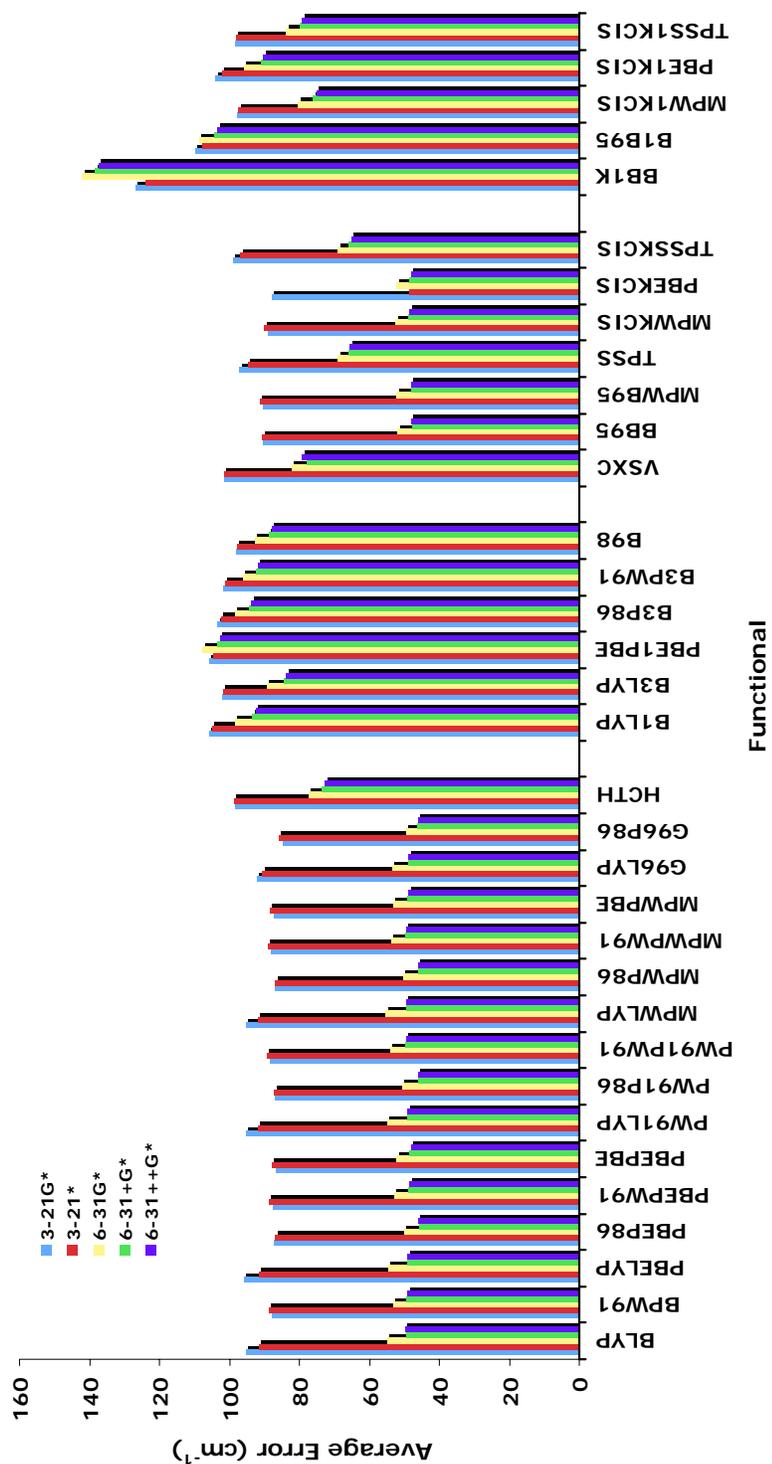


Figure 5-5. Average unsigned vibrational frequency errors (in cm^{-1}) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Pople type basis sets.

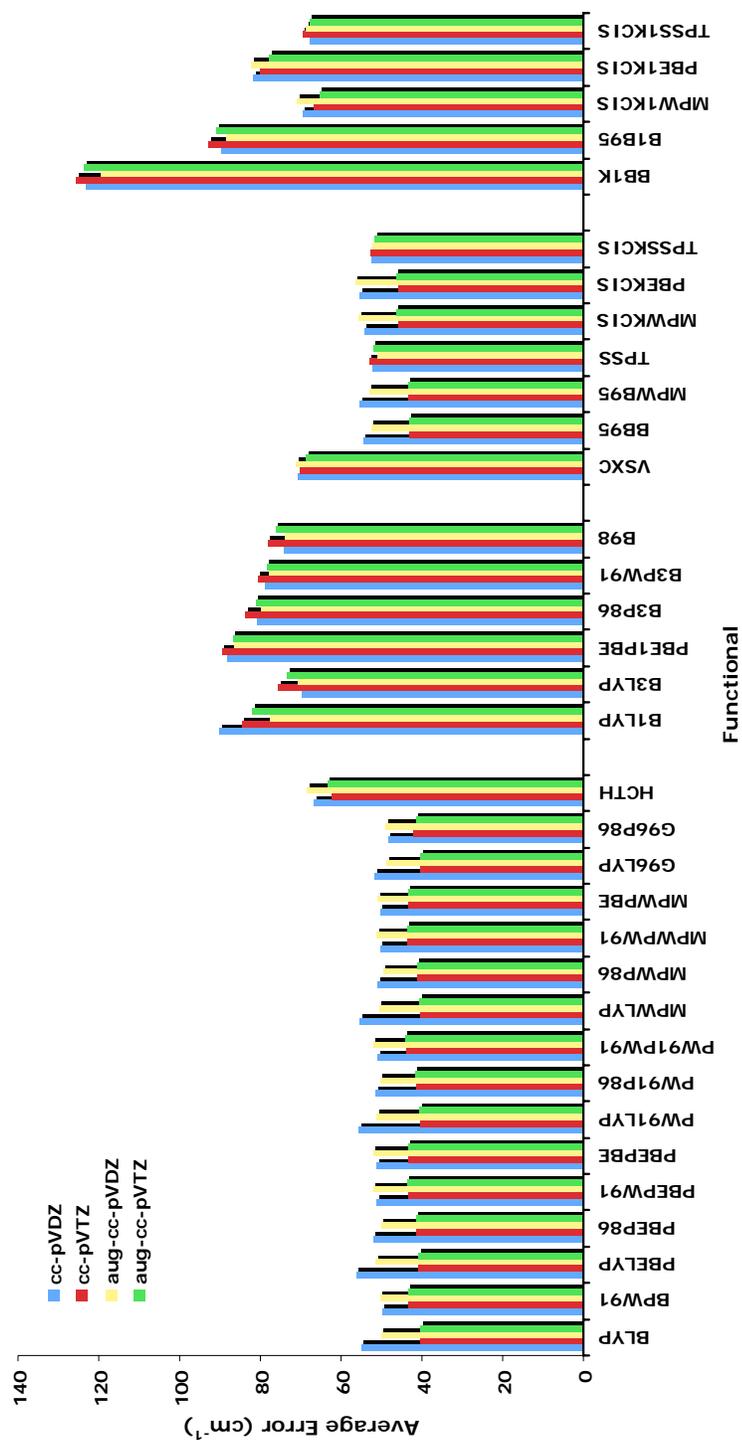


Figure 5-6. Average unsigned vibrational frequency errors (in cm^{-1}) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Dunning type basis sets.

Of the three LSDA functionals the best ionization potential unsigned error of 7.46 kcal/mol is obtained using the SPL/cc-pVDZ method. Among the GGA functionals G96P86/cc-pVTZ yields the lowest unsigned error of 4.93 kcal/mol, it should also be noted that the PW91PW91/cc-pVTZ, PBEP86/cc-pVTZ, PBEPW91/aug-cc-pVTZ, and PBEPBE/aug-cc-pVTZ methods all yield errors lower than 5.00 kcal/mol. The B98/aug-cc-pVTZ method yields an average unsigned error of 4.65 kcal/mol, the lowest unsigned error among all hybrid-GGA methods. The functional/basis set combination yielding the best result in the meta-GGA class of functionals is MPWB95/aug-cc-pVTZ with a calculated value of 4.38 kcal/mol. Among the hybrid-meta-GGA functionals the lowest average unsigned error is given by the B1B95/aug-cc-pVTZ method with a value of 4.25 kcal/mol.

As in the case of electron affinities, HF does a very poor job in predicting ionization potentials, this can be explained in the same way as above. The cation has fewer electrons than the neutral systems and, thus, exhibits less correlation effects. The Hartree-Fock method's inability to describe electron correlation leads to a more accurate prediction for the electronic energy of the cation as compared to the neutral species.

As one might expect, the small, 3-21G* and 3-21+G*, basis sets typically perform very poorly in predicting ionization potentials compared to the larger basis sets. Generally the 3-21+G* basis set predicts average errors that are substantially lower than the 3-21G* basis set. The best ionization potential result for the 3-21G* basis set is obtained with the hybrid-GGA B3LYP functional with an average error of 7.33 kcal/mol while the 3-21+G* basis set has its lowest average unsigned error of 5.26 kcal/mol when used in conjunction with the meta-GGA BB95 functional.

Inspection of the average unsigned errors for individual functionals in Figures 5-7 and 5-8 and in Table 5-15 reveals that the cc-pVTZ, 6-31+G*, 6-31++G*, aug-cc-pVDZ, and aug-cc-pVTZ basis sets all yield fairly similar results that are typically superior to the results obtained with the 3-21G*, 3-21+G*, 6-31G*, and cc-pVDZ basis sets. An exception to this is the LSDA functionals, for which the 6-31G* and cc-pVDZ basis sets yield the lowest average unsigned errors. As the Pople type 6-31+G* and 6-31++G* basis sets are computationally much less expensive to use compared to the larger Dunning correlation consistent basis sets, it is very promising, in terms of biological applications, that such high quality results can be obtained using the smaller basis sets. It should be noted that the MPWB95/(6-31+G*, 6-31++G*) methods (4.53 kcal/mol, 4.50 kcal/mol) outperform all other meta-GGA methods with the exception of the MPWB95/(cc-pVTZ, aug-cc-pVTZ) methods (4.49 kcal/mol, 4.38 kcal/mol). Similarly, the B1B95/(6-31+G*, 6-31++G*) methods (4.81 kcal/mol, 4.80 kcal/mol) yield better results than all other hybrid-meta-GGA methods with the exception of B1B95/(aug-cc-pVDZ, cc-pVTZ, aug-cc-pVTZ) methods (4.64 kcal/mol, 4.53 kcal/mol, 4.25 kcal/mol). For the 6-31++G* basis set, the best GGA result of 5.07 kcal/mol is obtained with the PBEPW91 functional. The best hybrid-GGA result for Pople type basis sets is 5.05 kcal/mol and is obtained with the B98/6-31+G* method.

As with the electron affinities, the addition of diffuse functions to hydrogen atoms in the 6-31++G* basis set seems to have a negligible effect on the calculation of ionization potentials compared to the 6-31+G* basis set. There is also only a small difference between the results obtained with the cc-pVTZ and aug-cc-pVTZ basis sets.

There is, however, a marked difference in the quality of the cc-pVDZ and aug-cc-pVDZ basis set results.

Table 5-15. Average unsigned ionization potential errors for HF, MP2, LSDA, and B3P86 methods.

Method	3-21G*	3-21+G*	6-31G*	6-31+G*	6-31++G*	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ	aug-cc-pVTZ
HF	26.53	23.00	25.70	24.74	24.76	26.33	25.74	25.19	25.40
MP2	15.59	9.77	10.45	10.42	5.84	11.44	9.78	6.74	5.53
SVWNV	10.74	10.29	7.57	8.52	8.48	7.68	8.33	8.59	8.75
SPL	10.50	10.37	7.49	8.62	8.57	7.46	8.42	8.70	8.83
cSVWNV	24.93	17.31	22.57	19.08	19.05	23.37	19.85	19.06	18.69
B3P86	15.05	19.01	14.25	15.87	15.86	13.83	15.22	15.84	15.68

Energies are in kcal/mol

Electron Affinity

The electron affinity average unsigned errors are shown in Table 5-16 and in Figure 5-9, the first of these gives the results for the HF, MP2, B3P86, and LSDA functional methods, the second gives all GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functional method results except for the hybrid-GGA functional B3P86, which yields very poor results. The best overall result for electron affinities is obtained using the meta-GGA functional MPWB95 along with the 6-31++G* basis set yielding an average unsigned error of 3.08 kcal/mol. The worst average unsigned error among the DFT methods is 16.78 kcal/mol and is obtained with the hybrid-GGA B3P86 functional combined with the 3-21+G* basis set. Among all methods studied in this work, the HF/aug-cc-pVTZ method yields the worst result with an average error of 29.63 kcal/mol.

Among all of the functional types considered here, the LSDA methods produce the highest errors; this is an expected result as these are the least sophisticated functionals and they lack gradient dependent terms. It is also interesting to note that results obtained with methods that incorporate the P86 correlation functional, with the exception of

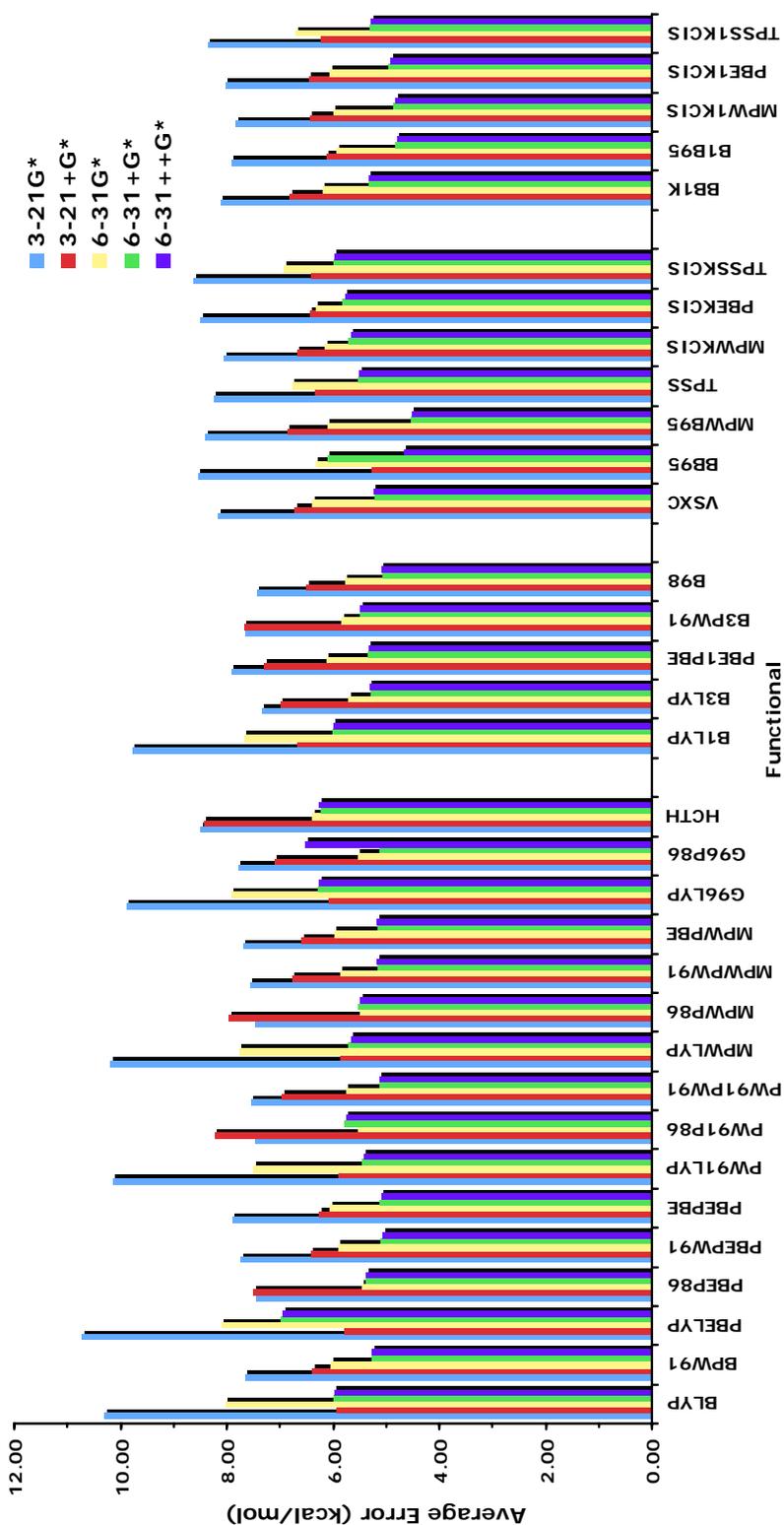


Figure 5-7. Average unsigned ionization potential errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals with Pople type basis sets.

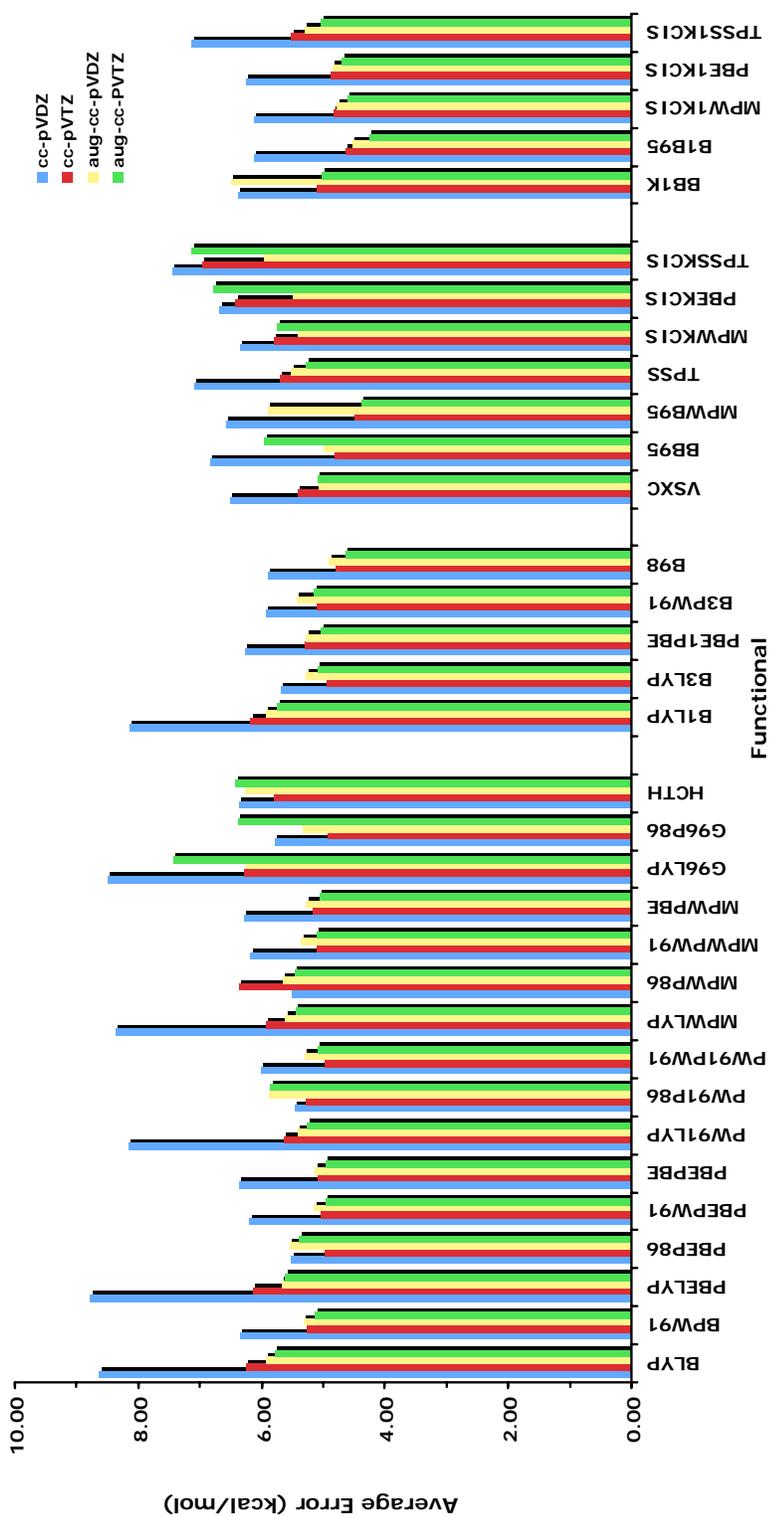


Figure 5-8. Average unsigned ionization potential errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals with Dunning type basis sets.

G96P86, are significantly worse than those obtained using the other functionals within a given class. Of the three LSDA functionals the best result of 6.25 kcal/mol is obtained with SVWN5/6-31+G*. Among the GGA functionals the PW91LYP/6-31++G* method yields the lowest average unsigned error of 3.56 kcal/mol, the PW91LYP/6-31+G*, PBEPW91/6-31++G*, and PBELYP/6-31++G* functional/basis set combinations all obtain errors lower than 3.60 kcal/mol. The B98/aug-cc-pVTZ method gives the smallest average error among the hybrid-GGA functionals of 3.15 kcal/mol. Among the meta-GGA functionals, the MPWB95/6-31++G* method yields the lowest error with a value of 3.08 kcal/mol, it is also noteworthy that the MPWB95/6-31+G* functional gives extremely good results with an error value of 3.12 kcal/mol. The MPW1KCIS/aug-cc-pVTZ method yields an error value of 3.48 kcal/mol, the lowest error among all hybrid-meta-GGA methods.

The Hartree-Fock method performs very poorly in describing electron affinities. This can be explained by the fact that, since an anion has one electron more than its neutral counterpart, correlation effects have a stronger effect on the negatively charged ion than on the neutral system. Due to the neglect of correlation effects in the Hartree-Fock technique, there is a pronounced discrepancy in its description of neutral and anionic species.

One salient aspect of these data is that, not surprisingly, the 3-21+G* basis set performs very poorly compared to the larger basis sets for all functionals. This basis set does, however, outperform all other basis sets when combined with the Hartree-Fock

method and gives results that are only slightly worse than those obtained with the larger, 6-31+G* and 6-31++G*, basis sets when combined with the MP2 method. The lowest unsigned error for the very small (and inexpensive) 3-21+G* basis set is 5.17 kcal/mol as calculated with the BB95 functional.

It is also surprising that, generally, the 6-31+G* and 6-31++G* basis sets obtain results that are comparable to, or in many cases superior to, the aug-cc-pVDZ and aug-cc-pVTZ results. As can be seen in Figure 7 and Table 3, the average unsigned errors for the 6-31+G* and 6-31++G* basis sets are generally lower than those for the aug-cc-pVDZ and aug-cc-pVTZ basis sets for the LSDA, GGA, and meta-GGA functional classes. This trend is especially pronounced for the LSDA and GGA functionals, two of the three LSDA functionals obtain better results when combined with the smaller, Pople 6-31G type, basis sets compared to the results they obtained when used in conjunction with the larger, Dunning cc-pVXZ type, basis sets. The Pople type basis sets outperform the Dunning type basis sets for thirteen of the sixteen GGA functionals. For the hybrid-GGA and hybrid-meta-GGA functionals the Dunning type basis sets typically outperform the Pople type basis sets by a small margin (≤ 0.5 kcal/mol). The meta-GGA functionals represent a “mixed bag” in terms of this trend, here the smaller basis sets outperform the larger ones (not including the 3-21+G* basis set) for four of the seven functionals studied in this work.

It should be noted that the addition of diffuse functions to hydrogen atoms in the 6-31++G* basis set does not lead to results that are significantly different than those obtained with the 6-31+G* basis set. It is also interesting to note that the aug-cc-pVTZ

basis set generally yields results that are only slightly better than the aug-cc-pVDZ basis set results.

Table 5-16. Average unsigned electron affinity errors for HF, MP2, LSDA, and B3P86 methods.

Method	3-21+G*	6-31+G*	6-31++G*	aug-cc-pVDZ	aug-cc-pVTZ
HF	26.40	29.05	28.95	28.90	29.63
MP2	11.14	10.73	10.62	5.81	4.70
SVWNV	9.37	6.25	6.36	7.58	7.53
SPL	9.92	6.79	6.90	8.14	8.04
cSVWNV	12.04	14.80	14.65	13.60	13.68
B3P86	16.78	13.02	13.08	13.99	13.53

Energies in kcal/mol

Heat of Formation

Unsigned errors for the heat of formation (HOF) test set are listed in Figures 5-10 and 5-11 along with Table 5-17. Overall, the combination that gives the lowest unsigned error is PBE1KCIS/aug-cc-pVTZ at 3.64 kcal/mol. Neglecting errors from the LSDA, MP2, and HF methods, the overall least accurate combination is PW91P86/6-31G* with an average unsigned error of 51.4 kcal/mol. The MPWLYP/3-21G* method yields an unsigned error of 5.66 kcal/mol which is the lowest error for the relatively inexpensive 3-21G* and 3-21+G* basis sets.

For the Pople basis sets, the accuracy of HOF calculations is dependent on the size of the basis set for the hybrid-GGA and meta-hybrid GGA classes of functionals. As shown in Figure 5-10, the all of the meta-hybrid GGA functionals and all but one of the hybrid-GGA functionals yield much higher errors for the 3-21G* and 3-21+G* basis sets than for the 6-31G*, 6-31+G*, and 6-31++G* basis sets while the other functional classes show no such dependency. The use of diffuse basis sets with GGA and meta-GGA

functionals appears to increase the accuracy of the methods as 3-21+G*, 6-31+G*, and 6-31++G* produce typically lower HOF errors than their non-diffuse counterparts. The

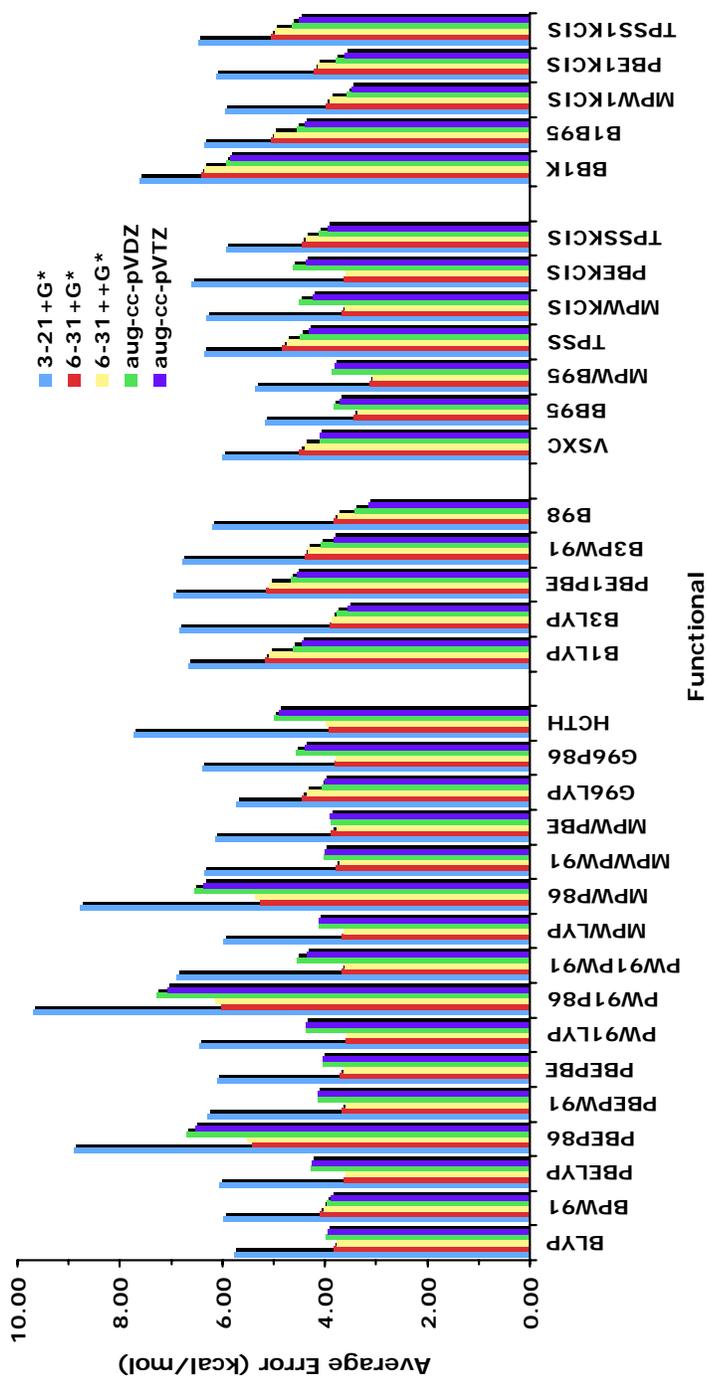


Figure 5-9. Average unsigned electron affinity errors (kcal/mol) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals.

opposite effect is observed when diffuse bases are paired with hybrid-GGA or hybrid-meta-GGA functionals. The most accurate functional/basis combination within the set of Pople bases is TPSSKCIS/6-31+G*, yielding an average unsigned error of 4.76 kcal/mol.

Generally, the hybrid-meta-GGA class of functionals produces the most accurate HOF calculations. Within this class B1B95/6-31G* yields the most accurate results with a 5.03 kcal/mol average unsigned error. For all five functionals in this class, the 6-31G* basis is the most accurate of the Pople type bases. The meta-GGA class of functionals yields errors slightly larger than those of the hybrid-meta-GGA class, on the whole. Functionals employing the TPSS exchange perform the best in this class. BB95 also performs well. TPSSTPSS/6-31+G* and TPSSKCIS/6-31+G* are the most accurate combinations within this functional category, producing average errors of 4.76 kcal/mol. The 6-31+G* basis yields the lowest error for each of the functionals in this class. Within the hybrid-GGA class, no functional's performance is particularly good. Within the GGA family, the P86 correlation term should be avoided as these data suggest that it is generally ill-suited for HOF calculations. Four functionals, PBELYP, MPWLYP, MPWPW91, and MPWPBE perform very well. With the exception of MPWP86, all functionals containing the MPW exchange perform well. MPWLYP/3-21G* gives the most accurate results of the entire class with an average unsigned error of 5.66 kcal/mol, which is remarkable for such an inexpensive method. The accuracy of this method surpasses that of many of the more expensive methods included in this study. As seen in Figure 5-10, the 6-31G* basis generally produces the highest average errors with a few exceptions.

As seen in Figure 5-11, the Dunning-style correlation-consistent basis sets yield a smaller range of errors than the Pople-style bases. This may be due to the fact that the cc-pVDZ basis, the smallest Dunning-style basis used, is still quite large and is more accurate than the 3-21G* and 3-21+G* bases. As discussed previously, enthalpy calculations were performed at the same functional/basis combination as the geometry optimization of each molecule, the exception being aug-cc-pVTZ, for which single point enthalpy calculations were performed at the TPSS1KCIS/aug-cc-pVDZ geometries. The lowest average error for all Dunning-style bases is obtained with B3PW91/aug-cc-pVTZ at 3.95 kcal/mol. Within each class of functional there is a mixture of accurate and inaccurate methods.

Within the hybrid-meta-GGA class, MPW1KCIS is the most accurate; with the MPW1KCIS/cc-pVTZ method producing an average error of 3.97 kcal/mol. Augmentation of the bases with diffuse functions tends to reduce the accuracy of methods in this functional class. Once again, the meta-GGA functionals prove to be the second most accurate DFT methods for HOF calculation. On the whole, the TPSSKCIS and TPSS methods produce the best results among meta-GGA functionals. However, VSXC/aug-cc-pVDZ is the most accurate combination in the class with an average unsigned error of 3.98 kcal/mol. The use of DZ versus TZ bases does not seem significant within this class, as the TZ bases produces the largest errors in three of the seven functionals. The hybrid-GGA class of functionals reveals the same trends with the Dunning-type bases as with the Pople bases. PBE1PBE is most accurate functional and B3PW91/aug-cc-pVTZ produces the lowest error at 3.95 kcal/mol. Within this class, the expansion of the basis set from DZ to TZ does not enhance the accuracy of the HOF

calculations as the TZ bases produce nearly equivalent errors for all functionals in the set. The behavior of the GGA class of functionals with the Dunning type basis sets is again similar to that of the Pople bases. Methods containing the P86 correlation term are again very poor at predicting heats of formation, while those containing the MPW exchange term are more accurate. BPW91 also performs well compared to the rest of the functionals in this group. The most accurate method within this class is HCTH/aug-cc-pVDZ, which produces an average error of 6.83 kcal/mol. Twelve of the sixteen functionals in the class show a decrease in accuracy with the addition of diffuse functions. Again, there is little difference in values obtained with DZ bases as opposed to TZ methods, as most functionals show only a slight increase in accuracy when using the TZ bases instead of the DZ sets.

Of the LSDA methods, c-SVWN5 performs notably well at predicting heats of formation. c-SVWN5/3-21+G* yields an average unsigned error of 9.44 kcal/mol, the best value in this group. In terms of heat of formation, c-SVWN5 is 4 to 13 times more accurate than other LSDA methods. Other LSDA methods do not accurately predict HOF.

Table 5-17. Average unsigned HOF errors for the HF, MP2, and LSDA methods.

Method	3-21G*	3-21+G*	6-31G*	6-31+G*	6-31++G*	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ	aug-cc-pVTZ
HF	289.94	294.01	252.94	256.28	258.10	261.27	248.30	259.40	248.49
MP2	58.23	55.95	83.74	84.53	83.27	96.10			
SVWNV	125.83	127.57	135.97	128.06	121.78	124.24	135.06	122.98	133.85
SPL	128.66	117.21	143.48	133.70	127.46	129.90	144.22	128.62	139.49
cSVWNV	13.45	9.44	19.47	26.46	27.14	19.48	29.44	18.95	18.10
Energies in kcal/mol									

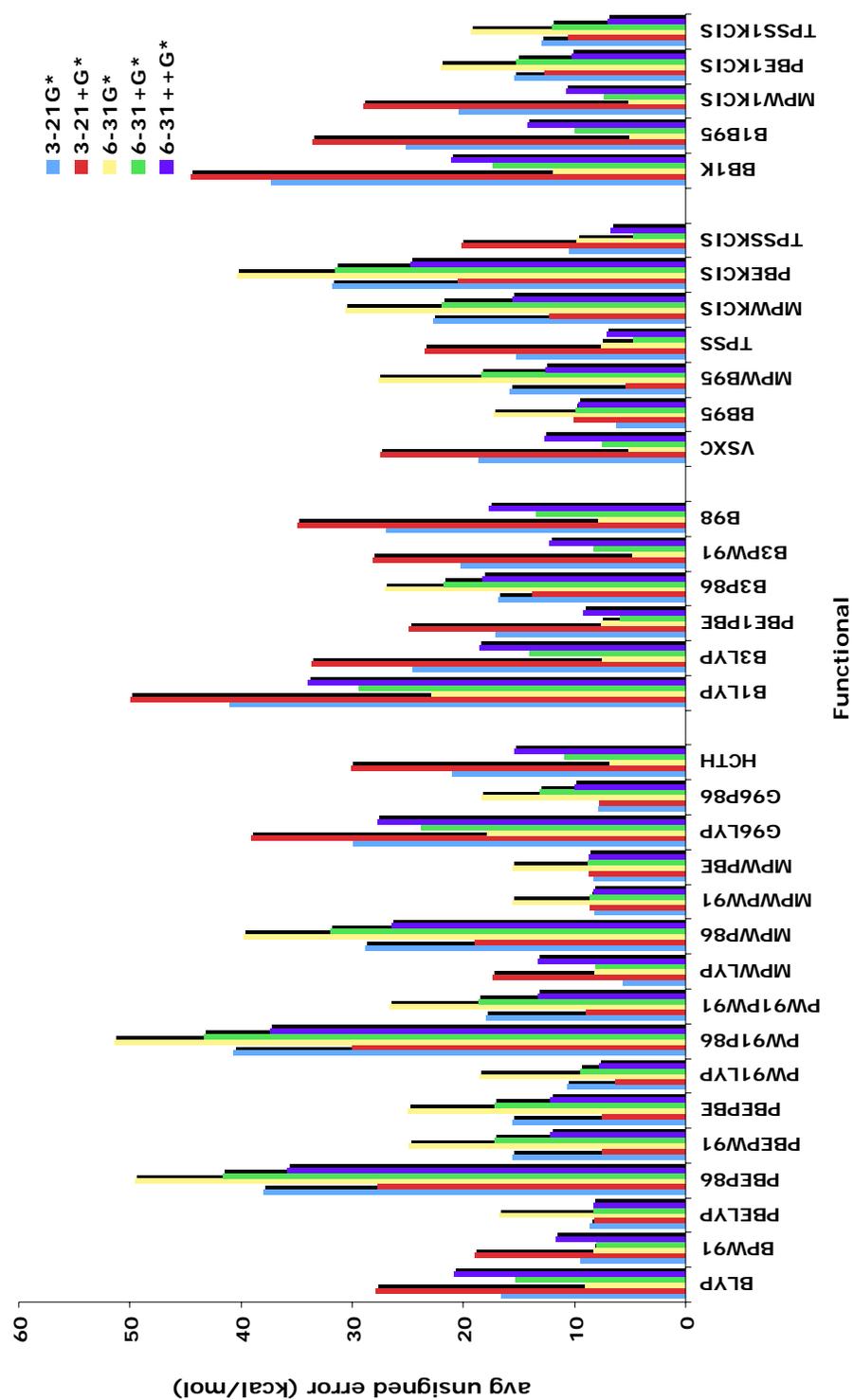


Figure 5-10. Average unsigned heat of formation errors (kcal/mol) for the five Pople-style basis sets employed in this study.

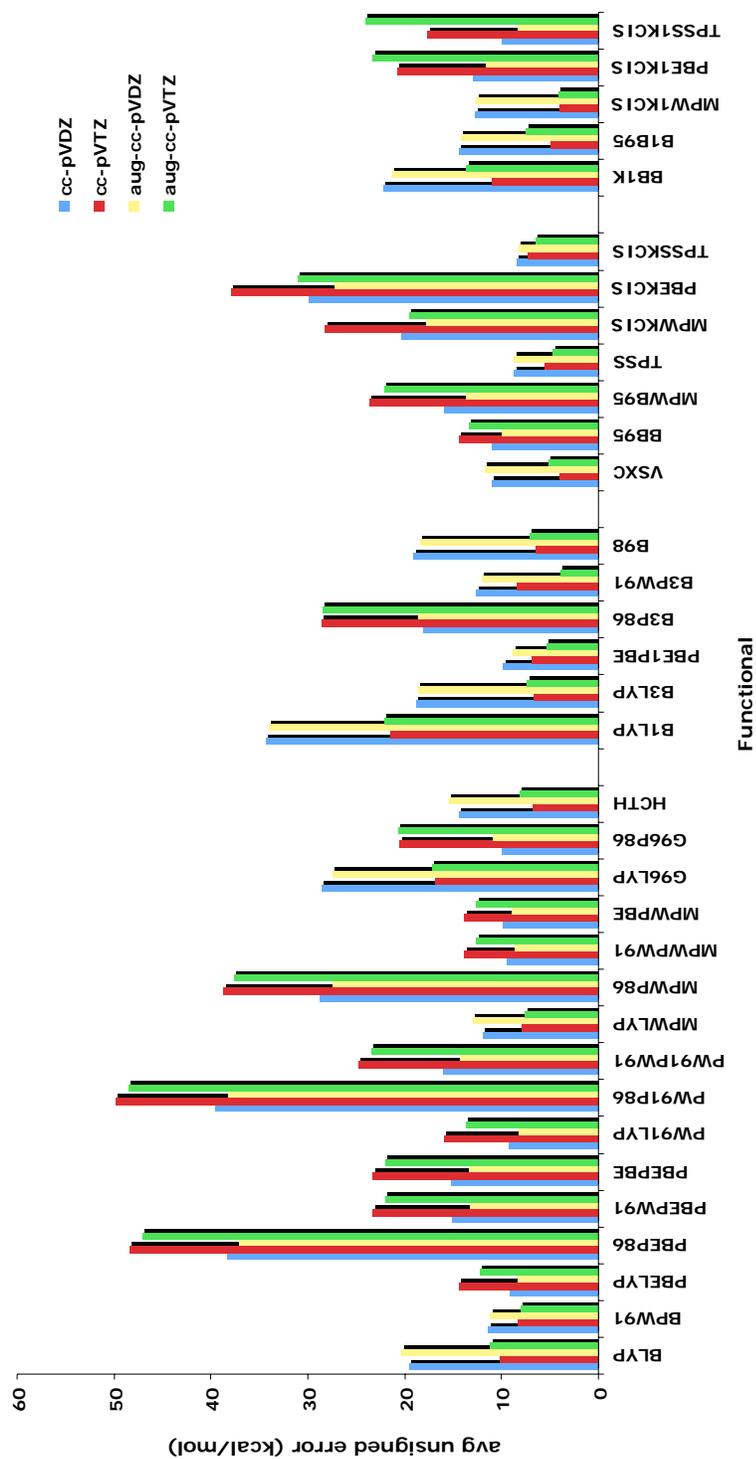


Figure 5-11. Average unsigned heat of formation errors (kcal/mol) for the Dunning-type basis functions used in this work.

Hydrogen Bonding Interaction Energy

Figures 5-12 and 5-13 give the average hydrogen bonding interaction energy unsigned errors for gradient corrected density functional methods along with Pople and Dunning type basis sets respectively. Table 5-18 gives the average unsigned hydrogen bond interaction errors for HF, MP2, and LSDA functional methods. Overall the best result is obtained with MP2/aug-cc-pVDZ with an average error of 0.25 kcal/mol. The best result among density functional methods is 0.31 kcal/mol as obtained by MPWLYP/aug-cc-pVTZ. The largest overall error of 10.26 kcal/mol is obtained by the c-SVWN5/aug-cc-pVDZ method.

Not surprisingly, the small Pople type basis sets, 3-21G* and 3-21+G*, generally yield poor results in terms of hydrogen bonding. For most functionals, the errors obtained with these small bases are greater than 2.00 kcal/mol. Some notable examples of small basis methods that perform fairly well are HCTH/3-21+G* (0.84 kcal/mol), G96LYP/3-21+G* (0.90 kcal/mol), and MP2/3-21G* (0.85 kcal/mol). The best result for these small basis sets combined with one of the LSDA methods, which are very computationally inexpensive, is 7.99 kcal/mol as calculated using c-SVWN5/3-21G*. As one might expect, the 6-31+G* and 6-31++G* basis sets, which contain diffuse functions, generally outperform 6-31G* in terms of hydrogen bonding, there are seven methods for which this is not the case, these are HF, BPW91, G96LYP, G96P86, VSXC, BB95, and B1B95. Somewhat surprisingly there is typically only a small advantage to using the 6-31++G* basis set, which incorporates diffuse functions for hydrogen atoms, as compared to the 6-31+G* basis set.

For the large Pople type basis sets, the MP2 method performs very well with average unsigned binding energies of 0.28 kcal/mol and 0.29 kcal/mol with 6-31++G*

and 6-31+G* respectively (these values represent the second and third best overall results). Hartree-Fock performs fairly well with these basis sets with a best value of 0.91 kcal/mol when combined with 6-31G*. The LSDA functionals perform poorly for hydrogen bonding when combined with the large Pople type basis sets. The SVWN5 and SPL functionals both yield errors greater than 6.00 kcal/mol with these bases. The c-SVWN5 functional, which gives results that are substantially better than those of the other two LSDA methods, still only yields a best result of 4.85 kcal/mol (with both the 6-31+G* and 6-31++G* basis sets).

There is a great deal of variation in the hydrogen bonding results obtained with the GGA functionals. The lowest interaction energy error of 0.46 kcal/mol is obtained with the MPWPW91 functional combined with both the 6-31+G* and 6-31++G* basis sets. The highest error of 2.59 kcal/mol is given by PW91P86/6-31G*. Other noteworthy methods in this class are MPWPBE/(6-31+G*,6-31++G*) (0.47 kcal/mol) and BLYP/6-31++G* (0.55 kcal/mol). It is interesting to note that, generally, functionals containing the P86 correlation functional perform poorly while functionals containing the MPW exchange functional perform fairly well when used along with the large Pople type basis sets. The MPWP86 functional performs moderately well with an average error of 1.03 kcal/mol for MPWP86/6-31++G*. For the large Pople type basis sets, the best result among hybrid-GGA methods is 0.33 kcal/mol as calculated with the B1LYP/6-31++G* method, it should also be noted that this is the best overall result for these basis sets among density functional methods. B1LYP/6-31+G* gives a slightly higher average unsigned interaction energy of 0.34 kcal/mol while B3LYP also performs well with average errors of 0.36 kcal/mol and 0.38 kcal/mol with 6-31+G* and 6-31++G*

respectively. Among the meta-GGA methods the lowest interaction energy error of 0.42 kcal/mol is obtained with the TPSS1KCIS/6-31++G* method. The VSXC functional performs very poorly compared to the other meta-GGA functionals (indeed, it performs poorly compared to most gradient corrected functionals). It is interesting to note that four of the seven functionals in this class obtain errors lower than 0.50 kcal/mol when combined with the 6-31+G* and 6-31++G* basis sets, these functionals are MPWB95, TPSS, MPWKICIS, and TPSSKICIS. Each of the five hybrid-meta-GGA functionals performs quite well for hydrogen bond interaction energies when paired with the 6-31+G* and 6-31++G* basis sets, with no method obtaining average errors larger than 1.00 kcal/mol. The best result in this class is 0.38 kcal/mol and is given by the MPW1KCIS/6-31++G* method. Other noteworthy methods are BB1K/6-31+G* (0.40 kcal/mol), BB1K/6-31++G* (0.41 kcal/mol), MPW1KCIS/6-31++G* (0.42 kcal/mol), and TPSS1KCIS/6-31+G* (0.42 kcal/mol).

As in the case of the Pople-type basis sets, the Dunning-type basis sets that contain diffuse functions, aug-cc-pVDZ and aug-cc-pVTZ, yield better hydrogen bond interaction energies than the ones that do not, cc-pVDZ and cc-pVTZ, for a majority of the functionals considered in this work. Generally speaking the cc-pVTZ functional outperforms the, smaller, cc-pVDZ basis set for hydrogen bonding, it should be noted that this is not the case for Hartree-Fock or any of the LSDA functionals. The aug-cc-pVTZ typically outperforms the aug-cc-pVDZ basis set for LSDA, GGA, and hybrid-GGA functionals while the smaller basis, aug-cc-pVDZ, yields better results when combined with the meta-GGA and hybrid-meta-GGA functionals.

HF yields fairly large errors with the Dunning-type basis sets, with the lowest unsigned error being 1.65 kcal/mol for HF/cc-pVDZ and the highest being 1.77 kcal/mol for HF/cc-pVTZ. These values are significantly higher than those obtained with the large Pople type basis sets. The MP2 method performs very well with most Dunning type basis sets, MP2/aug-cc-pVDZ produces an average unsigned error of 0.25 kcal/mol, which is the best value obtained for hydrogen bond interaction energies obtained in this work. Once again the LSDA functionals perform very poorly compared to the other DFT methods. Among these methods, the SPL and SVWN5 functionals generally yield results that are almost identical for all Dunning-type basis sets. The c-SVWN5 method yields unsigned errors that are significantly higher than those of SPL and SVWN5. The best LSDA result of 5.48 kcal is obtained with SPL/cc-pVDZ. The worst LSDA result is given by c-SVWN5/aug-cc-pVDZ with a value of 10.26 kcal/mol.

Among the GGA methods, the Dunning-type basis sets outperform the large Pople type bases for nine of the sixteen functionals. The MPWLYP functional performs significantly better than all other functionals in this class with the best value of 0.31 kcal/mol given by MPWLYP/aug-cc-pVTZ. The highest unsigned error of 2.78 kcal/mol is obtained with the G96LYP/cc-pVDZ method. Both functionals containing the G96 correlation functional, G96LYP and G96P86, give very poor results with errors that are greater than 2.00 kcal/mol for all Dunning-type basis sets.

For the hybrid-GGA methods, four of the six functionals, B1LYP, B3LYP, PBE1PBE, and B3P86, obtain errors that are between 0.50 kcal/mol and 0.80 kcal/mol, the remaining functional, B3PW91, does not perform as well, producing errors that are above 1.00 kcal/mol for all basis sets. For all Dunning-type basis sets, the B98 functional

produces the best hydrogen bonding results. The best result within this class is given by the B98/aug-cc-pVDZ method, with a value of 0.40 kcal/mol. Among the meta-GGA methods, PBEKCIS stands out as being notably better than all other functionals. Values of 0.32 kcal/mol and 0.34 kcal/mol are obtained when PBEKCIS is combined with the aug-cc-pVDZ and aug-cc-pVTZ basis sets respectively; these are the lowest unsigned errors within this class. The VSXC and BB95 functionals both perform poorly. Within the hybrid-meta-GGA class of functionals, PBE1KCIS yields the lowest errors for hydrogen bonding interaction energies for all Dunning-type basis sets. The best result in this class is obtained with PBE1KCIS/aug-cc-pVDZ with a value of 0.36 kcal/mol. The next best functional in this class is TPSS1KCIS, whose lowest error is 0.55 kcal/mol at the TPSS1KCIS/aug-cc-pVDZ level.

Table 5-18. Average unsigned hydrogen bond interaction errors for HF, MP2, and LSDA methods.

Method	3-21G*	3-21+G*	6-31G*	6-31+G*	6-31++G*	cc- pVDZ	cc- pVTZ	aug-cc- pVDZ	aug-cc- pVTZ
HF	1.13	1.60	0.91	1.07	1.08	1.56	1.77	1.68	1.73
MP2	0.85	1.14	0.49	0.29	0.28	1.29	0.42	0.25	0.30
SVWNV	9.68	10.24	6.48	6.20	6.21	5.49	5.76	6.04	5.97
SPL	9.66	10.24	6.48	6.21	6.21	5.48	5.76	6.04	5.97
cSVWNV	7.99	8.66	5.13	4.85	4.85	9.90	10.06	10.26	10.18

Energies in kcal/mol

Conformational Energies

The average unsigned conformational energy errors are given in Figures 5-14 and 5-15 and in Table 5-19. There are great differences in the conformational energies of the systems considered here. For example, the experimental difference in energy between the orthogonal and planar conformers of ethylene is 65.0 kcal/mol whereas the experimental value for the conformational energy for the anti and eclipsed forms of methanol is 1.1 kcal/mol.

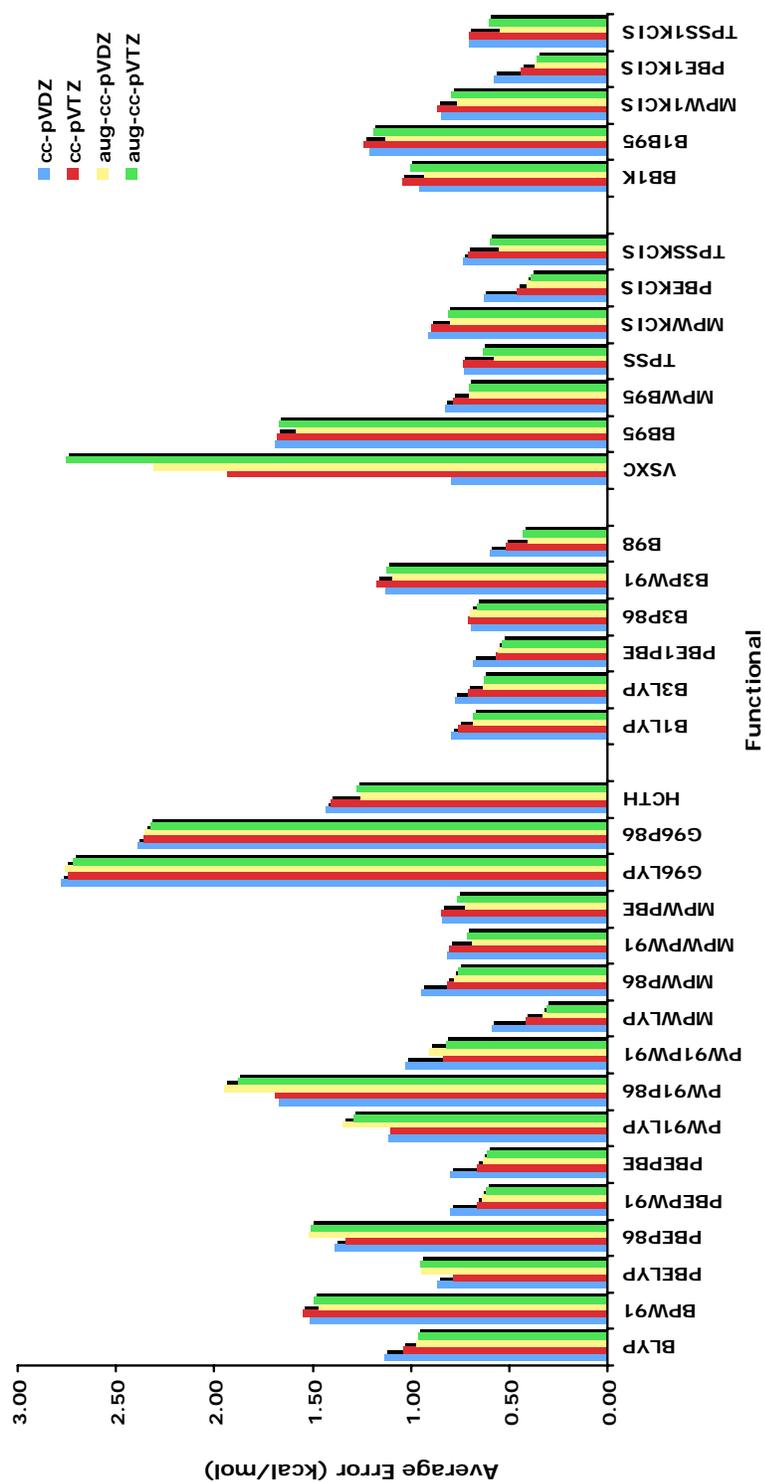


Figure 5-13. Average unsigned hydrogen bond interaction energy errors (kcal/mol) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Dunning type basis sets.

For this reason, the conformational energies are reported in percent error, that is:

$$error_{\%} = \frac{\Delta E_{\text{exp}} - \Delta E_{\text{theory}}}{\Delta E_{\text{exp}}} \times 100 \quad (5.24)$$

Overall, the best result of 6.8% is obtained with the MP2/aug-cc-pVDZ method.

The best result among density functional methods is 7.9% as calculated using MPWB95/cc-pVTZ. The worst conformational energy error is that of VSXC/3-21G* with a value of 81.9%.

As seen in Figure 5-14 and Table 5-19, the small Pople type basis sets, 3-21G* and 3-21+G*, give conformational energy errors that are typically much greater than those of the larger Pople type basis sets, 6-31G*, 6-31+G*, and 6-31++G*. Generally, 3-21+G* outperforms 3-21G*, there are several exceptions to this rule in the hybrid-GGA and hybrid-meta-GGA classes of functionals, also 3-21G* yields slightly lower errors than 3-21+G* for the GGA functional HCTH. For these small basis sets, the LSDA method produces conformational energies that are significantly worse than those of the gradient corrected density functional methods. The lowest unsigned error for small Pople type basis sets is obtained with the MP2/3-21+G* method with a value of 21.7%, for DFT methods the best value of 23.3% is obtained with the PBELYP/3-21+G* method.

For the Pople type basis sets the best conformational energy results can be found within the meta-GGA and hybrid-meta-GGA functional classes. The best overall result of 12.2% is obtained with the, hybrid-meta-GGA, B1B95/6-31++G* method. It should be noted that although the BB1K and B1B95 methods perform very well, the remaining three functionals in the hybrid-meta-GGA class, MPW1KCIS, PBE1KCIS, and TPSS1KCIS, yield errors that are about two to four percent higher. Within the meta-GGA

group of functionals, BB95, MPWB95, and TPSS all yield very low conformational energy errors. The lowest unsigned error in this class is produced by the MPWB95/6-31++G* method with a value of 12.4%. Among the hybrid-GGA functionals, B98 obtains errors that are about one percent lower than those of the next best functional, B3P86. The lowest error in this class is obtained at the B98/6-31++G* level with an average unsigned error of 14.2%. Two GGA functionals, PBEP86 and PW91P86, produce the best results within their class, both yield an error value of 14.0% when paired with the 6-31++G* basis set. Among the LSDA functionals, SPL and SVNWV both yield the same error values of 15.6% and 15.5% when paired with the 6-31+G* and 6-31++G* bases respectively. Hartree-Fock generates errors that are significantly higher than those obtained by most DFT methods, the best error value of 22.1% is obtained with 6-31+G*. The MP2 method obtains errors of 15.5% when paired with 6-31++G* and 16.8% with the 6-31+G* basis set.

The basis sets that include diffuse functions, 6-31+G* and 6-31++G*, generally give unsigned errors that are substantially lower than those obtained using the 6-31G* basis function. It is also interesting to note that 6-31++G* outperforms 6-31+G* for most of the functionals considered here. Figure 5-15 and Table 5-19 give the conformational energy unsigned errors for the Dunning type basis sets. Here it can be seen that there is no class of functional that stands out as being substantially more accurate than another. Some of the lowest unsigned errors are obtained with BB1K and B1B95 (hybrid-meta-GGA), BB95 and MPWB95 (meta-GGA), and with PBEP86 and PW91P86 (GGA). The B98 functional produces the best hybrid-GGA results, which are not quite as good as the best results obtained by other DFT methods. It is also interesting to note that each of the

LSDA methods studied here yields results that are competitive with many of those obtained with the, more sophisticated, gradient corrected techniques. Among all density functional methods considered in this work, the lowest unsigned error obtained for this property is 7.9% as calculated using the MPWB95/cc-pVTZ method. Once again, the VSXC functional (meta-GGA) performs very poorly for describing conformational energies. The MP2 method also yields very good results for all Dunning type basis sets except for cc-pVDZ. The best overall conformational energy result obtained in this study is 6.8% and is given by the MP2/aug-cc-pVDZ method. Hartree-Fock produces errors that are significantly higher than those of most DFT techniques.

Among the Dunning type basis sets, aug-cc-pVDZ and cc-pVTZ tend to yield the lowest errors. The aug-cc-pVDZ basis set gives the best results for all of the hybrid-GGA functionals, all of the LSDA functionals, and for all of the hybrid-meta-GGA functionals except MPW1KCIS. The cc-pVTZ basis set yields the lowest unsigned errors for all of the GGA functionals and for all of the meta-GGA functionals except VSXC and MPWKCIS. The cc-pVDZ basis set produces the largest errors among Dunning type basis sets for each of the computational techniques employed in this study with the exception of VSXC.

Table 5-19. Average unsigned conformational energy errors for HF, MP2, and LSDA methods.

Method	3-21G*	3-21+G*	6-31G*	6-31+G*	6-31++G*	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ	aug-cc-pVTZ
HF	27.6	38.7	24.2	22.1	22.2	27.7	20.3	22.0	21.0
MP2	30.8	21.7	18.8	16.8	15.5	19.8	7.4	6.8	8.9
SVWNV	51.9	39.1	19.5	15.6	15.4	17.1	9.9	8.8	11.6
SPL	52.1	39.3	19.7	15.6	15.5	17.4	10.1	8.9	11.6
cSVWNV	51.2	34.4	17.4	17.0	17.0	17.8	10.2	10.0	15.6

Values in percent error

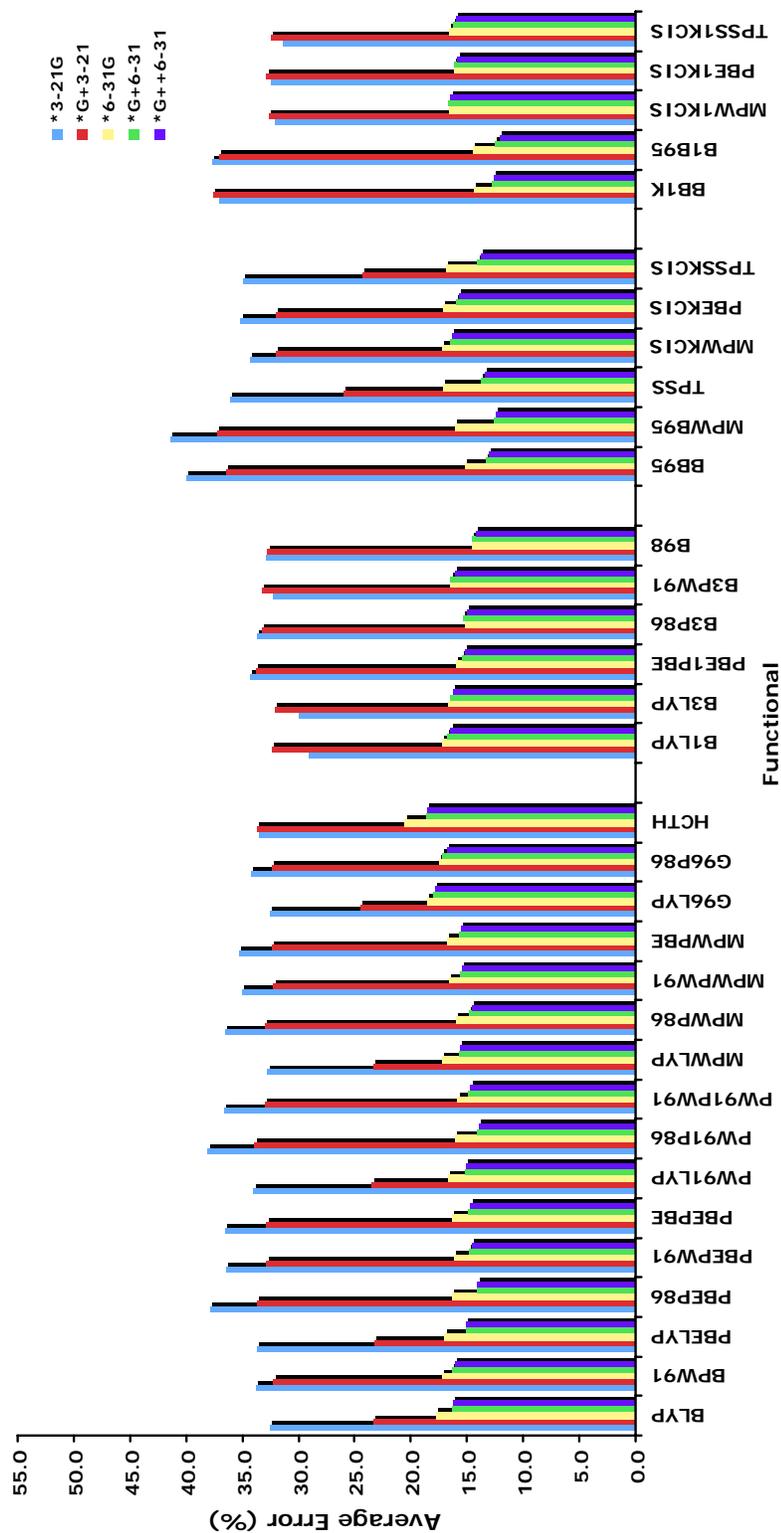


Figure 5-14. Average unsigned conformational energy errors (kcal/mol) for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Pople-type basis sets.

Reaction Barrier Heights for Small Systems with Non-singlet Transition States

Figures 5-16 and 5-17 give the average unsigned barrier height errors of the SRBH systems for gradient corrected functionals along with the Pople and Dunning type basis sets respectively. Table 5-20 gives the SRBH barrier height errors for the HF, MP2, and LSDA functional methods along with all basis sets considered in this work. Overall the best result is obtained with the B3LYP/aug-cc-pVTZ method with an average unsigned error of 1.05 kcal/mol. The highest error, 21.95 kcal/mol, is produced with the SVWN5/3-21G* functional/basis combination. Again we would like to point out that these barrier heights are based on single point calculations at geometries determined at the QCISD/MG3 level of theory.

Inspection of these data reveals that the DFT methods that include exact exchange, that is, the hybrid-GGA and hybrid-meta-GGA methods generally yield the lowest barrier height errors. The LSDA methods, which are based solely on the electron density, produce the largest unsigned errors. The LSDA, GGA, and meta-GGA perform poorly for SRBH barrier heights. Each of the LSDA methods produces errors larger than twelve kcal/mol for all basis sets. Of the GGA functionals, only HCTH yields errors smaller than six kcal/mol. The best result in this class is obtained with the HCTH/6-31++G* method with an average unsigned error of 4.86 kcal/mol. Among the meta-GGA methods, only the VSXC functional obtains errors smaller than six kcal/mol. The smallest error in this class is 4.24 kcal/mol and is given by the VSXC/6-31++G* method.

Among the hybrid-GGA functionals, B3LYP yields the smallest errors for all basis sets, this functional produces its lowest error of 3.11 kcal/mol when paired with the aug-cc-pVTZ basis set. It should be noted that B3LYP/6-31++G* gives a slightly higher error

of 3.23 kcal/mol. In the hybrid-meta-GGA class, the BB1K functional stands out as clearly being the best performer, indeed, for each basis set, this functional produces the best results among all methods considered in this work. The lowest error in this class is obtained with the BB1K/aug-cc-pVTZ method with a value of 1.05 kcal/mol. The next best functional for the calculation of these SRBH barrier heights is B1B95, which produces the second best results among all methods studied here. The lowest error given by this functional is 2.64 kcal/mol as calculated using the 6-31G* basis set.

The Hartree-Fock method performs very poorly in describing radical transition state barrier heights, the lowest unsigned error attained with this technique is 10.78 kcal/mol, with the 3-21+G* basis set. MP2 yields fairly good results when paired with the Dunning type basis sets but, when paired with the Pople type basis sets, produces much larger errors. The lowest unsigned error attained with this method is 2.98 kcal/mol at the MP2/aug-cc-pVDZ level.

Table 5-20. Average unsigned errors for the non-singlet transition state reaction test set for HF, MP2, and LSDA methods.

Method	3-21G*	3-21+G*	6-31G*	6-31+G*	6-31++G*	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ	aug-cc-pVTZ
HF	10.79	10.78	12.49	12.84	12.78	11.49	12.51	12.08	13.10
MP2	6.70	5.66	6.53	6.63	6.45	3.46	3.35	2.98	3.14
SVWNV	21.95	19.83	17.73	16.70	16.73	19.31	17.65	18.16	17.12
SPL	21.94	19.75	17.71	16.75	16.79	19.26	17.61	18.20	17.26
cSVWNV	18.82	16.38	14.18	14.18	13.23	16.05	14.18	14.70	13.46

Energies in kcal/mol

Reaction Barrier Heights for Organic Molecules with Singlet Transition States

Figures 5-18 and 5-19 and Table 5-21 show the reaction barrier heights for the six reactions listed in the LSBH test set. Transition state barrier heights in this study are calculated as the difference between temperature-corrected total enthalpy of the transition state and that of the reactants. All structures have been fully optimized at each functional/basis set combination.

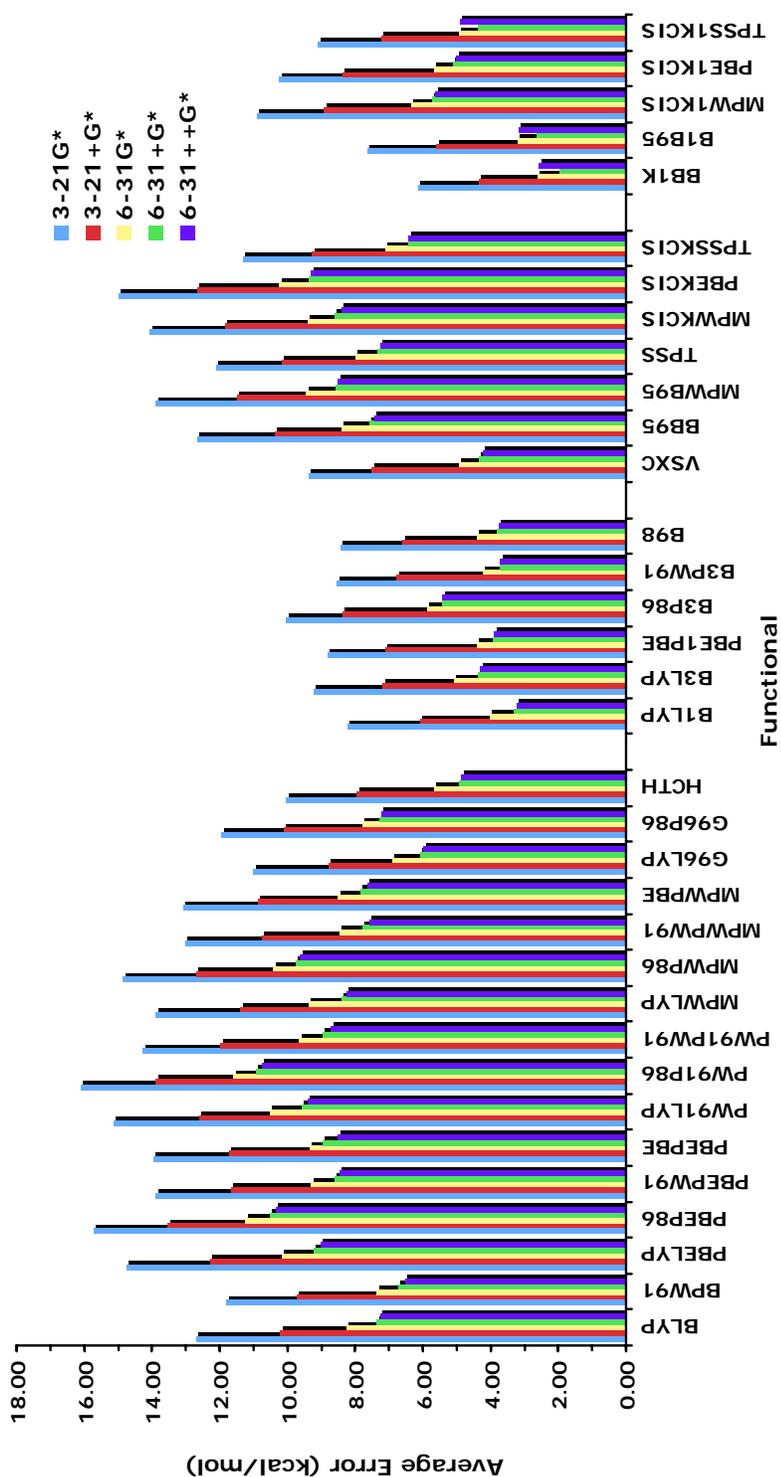


Figure 5-16. Average unsigned barrier height energy errors (kcal/mol) for SRBH reactions along with the GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Pople type basis sets.

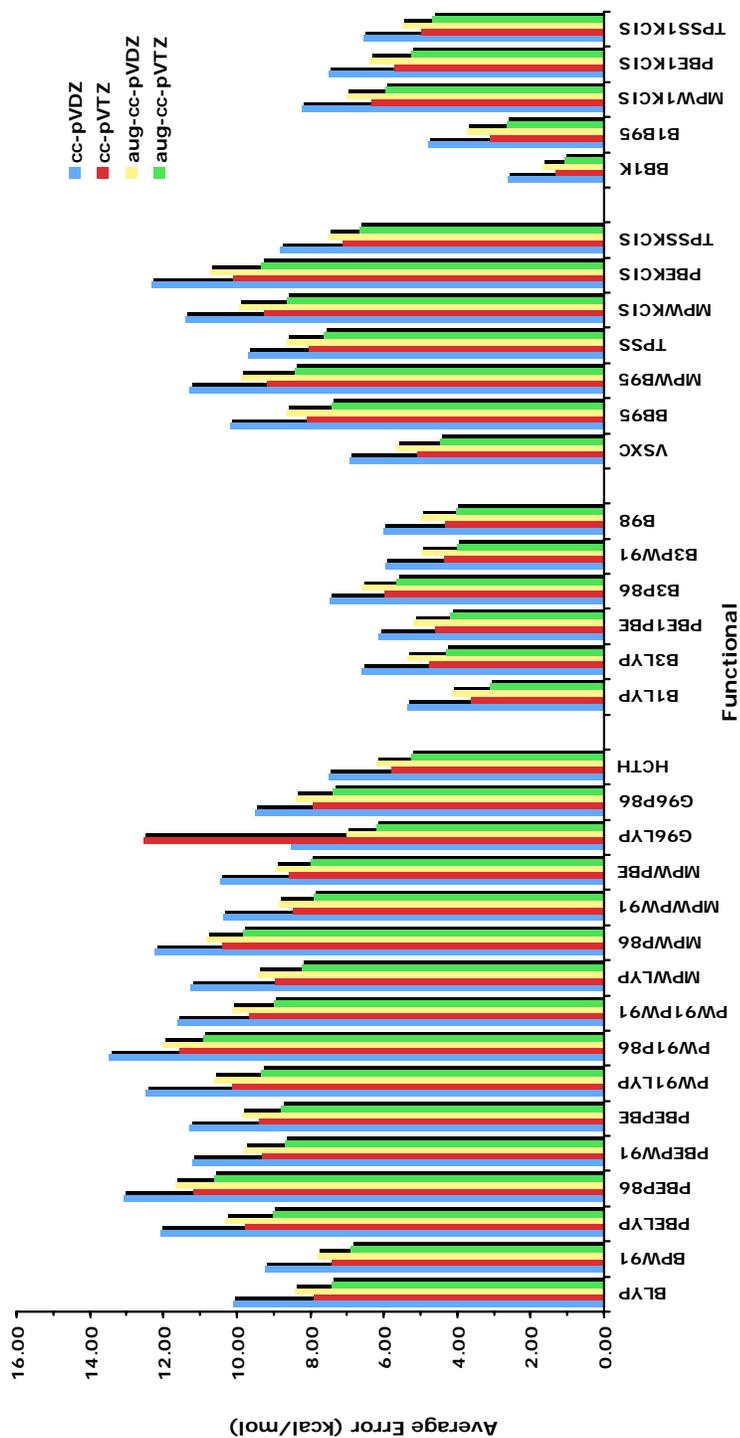


Figure 5-17. Average unsigned barrier height energy errors (kcal/mol) for SRBH reactions along with the GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Dunning-type basis sets.

The values listed in the following tables are average values of the error in transition state barrier height over all six reactions considered. These reactions include: 1) Diels-Alder reaction of butadiene and ethene forming cyclohexene 2) Cope rearrangement of 1,5 hexadiene 3) Claisen rearrangement of allyl vinyl ether to pentenal 4) the electrocyclic rearrangement of cyclobutene to butadiene 5) 1,5-sigmatropic shift of 2,4 pentanedione and 6) the 1,5-sigmatropic shift of 1,3-pentadiene.

Overall, the functional that provides the lowest average error over all six reactions for both Pople and Dunning basis sets is B1LYP. The average error for this functional is 2.63 kcal/mol for the 6-31++G* basis and 2.58 kcal/mol for the aug-cc-pVTZ basis. Generally, a marked improvement in accuracy is observed between basis sets for each functional. The 3-21G* and 3-21+G* are less accurate than the larger Pople-type bases by 3-4 kcal/mol, while the triple-zeta Dunning-style basis sets are more accurate than their double-zeta counterparts by nearly 0.5 kcal/mol. The hybrid-GGA and hybrid-meta GGA functional classes perform markedly better for predicting barrier heights than the LSDA, GGA, and meta-GGA classes. This result indicates a trend that is the opposite of that observed for frequency calculations, for which functionals that include the Hartree-Fock exact exchange perform worse than those without a DFT "exact exchange" term. Since frequency calculations must be performed for transition state optimizations this result is somewhat surprising. Moreover, on its own, the HF method is more accurate than most DFT methods at predicting barrier heights when the 3-21G* and 3-21+G* basis sets are used. MP2 also performs well with the lower basis sets. In fact, for HF, basis sets larger than 3-21G* produce errors nearly twice as large as those given by the smallest bases.

Among the LSDA functionals, c-SVWN5 functional gives the greatest accuracy, while SPL is slightly less accurate. The average barrier height error for c-SVWN5/cc-pVDZ is 11.22 kcal/mol. Typically, average errors within the LSDA class are near 12 kcal/mol except for the lower Pople style basis sets, which returned errors from 18-20 kcal/mol for the SPL and SVWN5 functionals. Within the GGA class, most functionals yield similar results, while the HCTH functional clearly returning the most accurate results. As mentioned, the accuracy of the barrier height calculations is highly basis set dependent for the Pople-type basis sets. HCTH/6-31+G* and HCTH/6-31++G* both yield an average error of 4.15 kcal/mol over all six reactions, while HCTH/cc-pVTZ produced an average error of 4.05 kcal/mol. For the class as a whole, errors for the 3-21G* and 3-21+G* average 9-13 kcal/mol while errors for the larger Pople bases average 6-8 kcal/mol. The Dunning basis sets provide accuracy equivalent to the high level Pople sets. Similar results are obtained by the meta-GGA class, in which VSXC is by far the most accurate. Once again the lower Pople basis sets yield average errors of 8-11 kcal/mol, while the larger Pople sets and the correlation-consistent sets give average errors of 7-8 kcal/mol. The VSXC functional consistently yields lower errors than the other functionals in this class.

Hybrid-GGA methods perform better than either the GGA or meta-GGA methods, with B1LYP proving to be the most accurate functional tested. Again a large dependence on the basis is observed with the Pople-type basis sets as the larger basis sets are much more accurate than 3-21G* and 3-21+G*. TZ Dunning-type sets are slightly more accurate than the DZ sets. B1LYP/6-31++G* and B1LYP/aug-cc-pVTZ are the most accurate functional/basis combinations in the entire test set, producing average errors of

2.63 and 2.58 kcal/mol, respectively. B3LYP also provides very accurate calculations for the barrier height test set. On the whole, the hybrid-GGA and hybrid-meta-GGA classes provide similar accuracy. Of the hybrid-meta-GGA class, BB1K yields the lowest average errors with the Pople-type bases, while B1B95 performs better when the Dunning-style sets are employed.

Table 5-21. Average unsigned errors for the singlet transition state large reaction test set for HF, MP2, and LSDA methods.

Method	3-21G*	3-21+G*	6-31G*	6-31+G*	6-31++G*	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ	aug-cc-pVTZ
HF	7.79	8.69	13.93	13.89	13.89	13.73	14.32	13.39	14.39
MP2	5.68	5.41	5.18	5.28	7.07	7.71	8.80		
SVWNV	19.86	17.81	12.61	12.04	12.08	12.01	11.83	12.54	12.21
SPL	19.82	17.78	12.59	12.02	12.05	11.99	11.75	12.49	11.83
cSVWNV	14.11	13.25	13.01	11.31	11.35	11.22	11.93	11.74	12.33

Energies in kcal/mol

Conclusions

In terms of geometric parameters hybrid-GGA and hybrid-meta-GGA generally yield the best results for both bond lengths and bond angles. The LSDA functionals generally do not perform as well as the more sophisticated functionals. The choice of basis set has a large impact on the quality of calculated geometric parameters. In terms of bond lengths, the large Pople type basis sets, 6-31G*, 6-31+G*, and 6-31++G*, generally perform similarly to or better than the much larger (and more expensive) cc-pVDZ and aug-cc-pVDZ basis sets for all gradient corrected functionals. For bond angles, the Dunning type basis sets generally yield the best results. The largest of these bases, aug-cc-pVQZ, generally obtains the lowest bond angle errors for all DFT functional classes. The large Pople type basis sets that incorporate diffuse functions typically yield bond angles that are only slightly less accurate than those obtained with the aug-cc-pVDZ and cc-pVTZ basis sets. For most functionals, 6-31++G* produces bond angle errors that are only 0.01° to 0.05° higher than those of aug-cc-pVDZ.

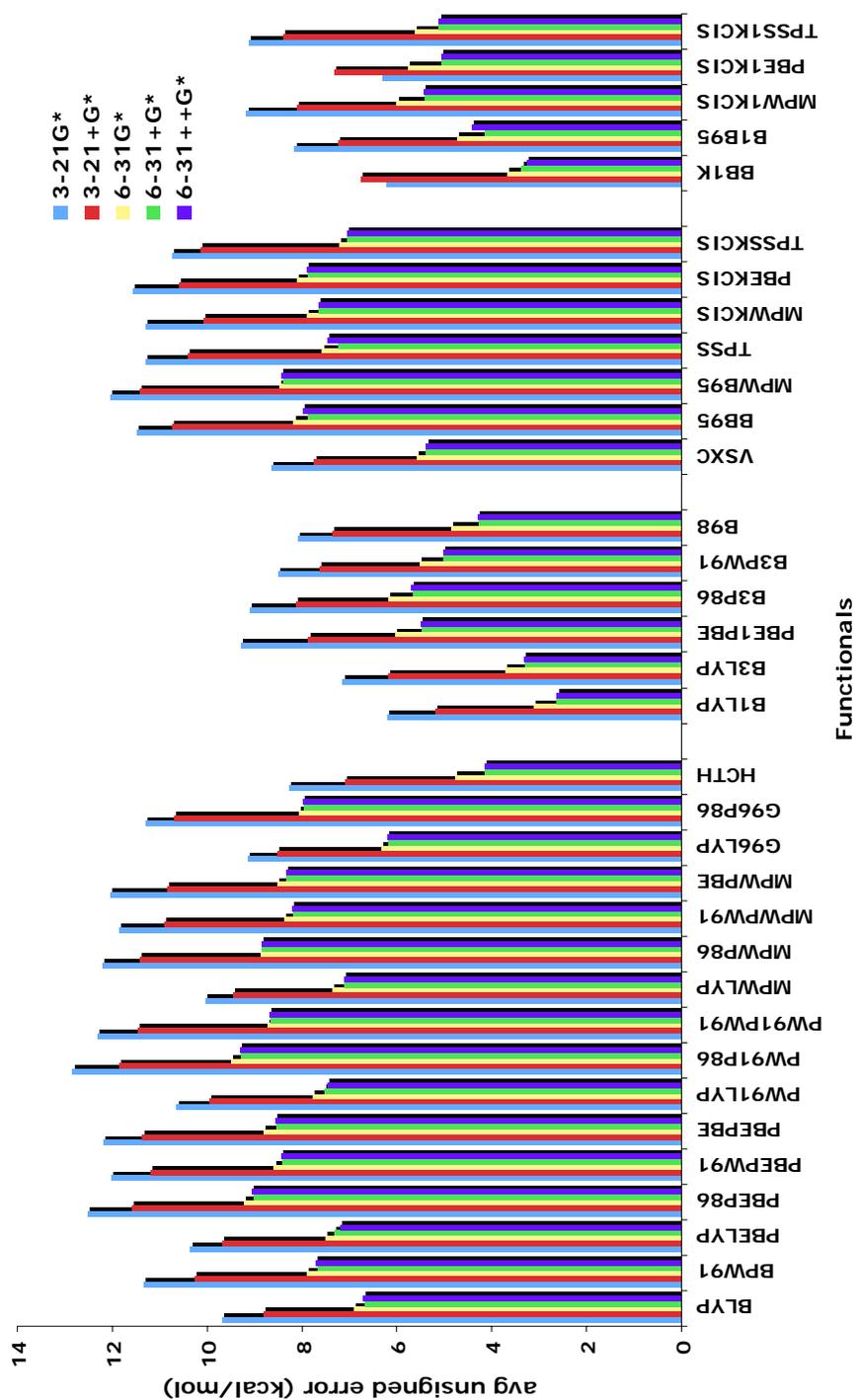


Figure 5-18. Average unsigned barrier height energy errors (kcal/mol) for large singlet transition state reactions along with the GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Pople-type basis sets.

The methods that include DFT “exact exchange” perform very poorly for calculating the vibrational frequencies of molecules. For large Pople and Dunning type basis sets, these methods generally yield unsigned frequency errors that are 1.5 to 2 times larger than those obtained with method that do not include exact exchange. For all basis sets, with the exception of 3-21G*, the GGA functionals produce the lowest average frequency errors. For LSDA and GGA functionals, the augmented Pople type basis sets, 6-31+G* and 6-31++G*, typically produce errors that are slightly lower than those of aug-cc-pVDZ and slightly higher than those of aug-cc-pVTZ. For all functionals, the Pople-type basis sets yield errors that are comparable to the errors computed using all Dunning-type basis sets.

For electron affinities there is no strong tendency for one functional class to significantly outperform another, with the exception of LSDA, which performs very poorly compared to all other functional groups. It is interesting to note that all functionals containing the P86 correlation functional (GGA and hybrid-GGA) perform very poorly. Functionals incorporating “exact exchange” tend to yield the smallest errors when combined with larger Dunning-type basis sets while the other functional groups, LSDA, GGA, and meta-GGA, all obtain the most accurate results when used in conjunction with the 6-31+G* and 6-31++G* basis sets.

For ionization potentials the best results are obtained with the hybrid-meta-GGA functionals. It is very promising, in terms of large-scale calculations, that the ionization potential results obtained with the 6-31+G* and 6-31++G* basis sets are comparable to those obtained using the much larger Dunning-type basis sets for most functionals. As

one might expect, the inclusion of diffuse functions in the basis set greatly improves the results for this property.

For heats of formation, the meta-GGA and hybrid-meta-GGA classes of DFT functionals appear to be the most accurate. It is important to note that in all classes except LSDA, one can find some functional/basis combination that performs well. Overall, the Dunning-style bases are more accurate than the Pople-type sets, with the cc-PVTZ and aug-cc-pVTZ bases yielding the lowest average unsigned errors for our 156-molecule heat of formation test set. However, it should be noted that one can achieve a very high level of accuracy with the MPWLYP/3-21G* method. This combination produces an average error of only 5.6 kcal/mol, which is only 2 kcal/mol less accurate than the best result obtained within the entire study. Within the GGA class of functionals, a wide range of accuracies is obtained.

Generally the hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals yield the best results for hydrogen bond interaction energies. There is a large amount of variation among the GGA functionals, with some giving very good results and others performing very poorly. The MP2 method produces some of the lowest hydrogen bonding interaction energy errors. For both the large Pople type basis sets and the Dunning type bases the addition of diffuse functionals typically produces lower unsigned errors. The inclusion of diffuse functions on hydrogen atoms in the 6-31++G* basis does not generally increase the performance in terms of hydrogen bonding interaction energies when compared to the 6-31+G* basis. For the large Pople and Dunning type bases that include diffuse functions, there is no clear tendency for one particular basis set to

consistently produce the lowest errors within the GGA class of functionals, for all of the other functional classes the 6-31+G* and 6-31++G* bases generally give the best results.

In terms of conformational energies, the meta-GGA and hybrid-meta-GGA functionals produce the lowest average errors. Not surprisingly, the large Pople type basis sets, 6-31G*, 6-31+G*, and 6-31++G*, yield results that are typically about ten percent better than those obtained using the smaller Pople type bases, 3-21G* and 3-21+G*. For the large Pople type basis sets there is a slight improvement in the calculated conformational energies when diffuse functionals are employed. Overall the basis sets that produce the lowest errors are the Dunning type bases, aug-cc-pVDZ and cc-pVTZ.

One of the most salient aspects of the data concerning the barrier heights of small molecules with radical transition states (SRBH) is that functionals containing exact exchange terms generally produce the lowest average barrier height errors. The LSDA methods, which depend only on the electron density, produce errors that are significantly higher than those of all other methods considered here. In terms of basis sets, the inclusion of diffuse functions typically increases the accuracy with which the barrier heights of these reactions can be calculated. The lowest barrier height errors are generally produced with the 6-31+G*, 6-31++G*, and aug-cc-pVTZ bases.

As in the case of the SRBH reactions, the barrier heights of larger systems with singlet transition states (LSBH) are generally better described by functionals that contain exact exchange. The addition of diffuse functions to the 3-21G*, 6-31G*, and cc-pVTZ basis sets generally results in a lower unsigned average error, in the case of the cc-pVDZ basis set however, the addition of diffuse functions typically increases the errors slightly.

For the LSBH reactions, the 6-31+G*, 6-31++G*, and aug-cc-pVTZ basis sets generally produce the lowest errors for most methods studied in this work.

General Summary of the Survey of DFT Methods

Here, we will attempt to summarize the results obtained in the entire study and draw some conclusions concerning the functionals that seem to offer the best compromise in terms of describing all of the physical properties investigated in this work. As we have generated a tremendous amount of data in this study, we will limit our discussion by considering only the results obtained by two popular basis sets, 6-31+G* and aug-cc-pVDZ.

One of the most interesting observations that can be made from the data presented here is that, for many physical properties, the large Pople-type basis sets (6-31G*, 6-31+G*, and 6-31++G*) produce results that are comparable to, or superior to, those given by the much larger and computationally expensive Dunning-type basis sets. For example, for the B1B95 functional, the 6-31+G* basis set outperforms the aug-cc-pVDZ basis set for bond distances, heats of formation, hydrogen bond interaction energies, and reactions barrier heights (both SRBH and LSBH); the average unsigned bond angle error obtained with the smaller basis set is only 0.034° higher than that of the larger basis and the average unsigned ionization potential error for 6-31+G* is only 0.28 kcal/mol larger than that of aug-cc-pVDZ. The average unsigned electron affinity, vibrational frequency and conformational energy errors are larger for 6-31+G* than for aug-cc-pVDZ.

Table 5-22 indicates the rankings of the top five functional/basis set combinations overall and the top three functional/basis set combinations among Pople-type basis sets for each property considered in this work. In the table, it can be seen that for each

Table 5-22. Rankings of functional/basis set combinations for properties considered in this work.

Rank		Avg. Unsigned Error
Bond Length		
1	VSXC/cc-pVQZ	0.0056
2	VSXC/aug-cc-pVQZ	0.0057
3	VSXC/aug-cc-pVTZ	0.0061
4	VSXC/cc-pVTZ	0.0061
5	TPSS1KCIS/cc-pVTZ	0.0063
1	B1B95/6-31+G*	0.0075
2	B1B95/6-31++G*	0.0075
3	B1B95/6-31G*	0.0078
Bond angle		
		(deg)
1	BLYP/aug-cc-pVTZ	1.07
2	PBE1PBE/aug-cc-pVQZ	1.11
3	B3P86/aug-cc-pVQZ	1.12
4	PBE1PBE/cc-pVQZ	1.12
5	B3P86/aug-cc-pVDZ	1.12
1	PBE1PBE/6-31++G*	1.22
2	PBE1PBE/6-31+G*	1.23
3	TPSSTPSS/6-31++G*	1.23
Frequencies		
		(cm ⁻¹)
1	G96LYP/aug-cc-pVTZ	40
2	PW91LYP/cc-pVTZ	40
3	BLYP/aug-cc-pVTZ	40
4	G96LYP/cc-pVTZ	40
5	MPWLYP/cc-pVTZ	40
1	PBEP86/6-31+G*	46
2	PBEP86/6-31++G*	46
3	MPWP86/6-31++G*	46
EA		
		(kcal/mol)
1	MPWB95/6-31++G*	3.08
2	MPWB95/6-31+G*	3.12
3	B98/aug-cc-pVTZ	3.15
4	BB95/6-31++G*	3.35
5	B98/aug-cc-pVDZ	3.42
1	MPWB95/6-31++G*	3.08
2	MPWB95/6-31+G*	3.12
3	BB95/6-31++G*	3.35

IP		(kcal/mol)
1	B1B95/aug-cc-pVTZ	4.25
2	MPWB95/aug-cc-pVTZ	4.38
3	MPWB95/cc-pVTZ	4.49
4	MPWB95/6-31++G*	4.50
5	MPWB95/6-31+G*	4.53
1	MPWB95/6-31++G*	4.50
2	MPWB95/6-31+G*	4.53
3	BB95/6-31++G*	4.67
HOF		(kcal/mol)
1	B3PW91/aug-cc-pVTZ	3.95
2	MPW1kcis/cc-pVTZ	3.97
3	VSXC/cc-pVTZ	3.99
4	MPW1KCIS/aug-cc-pVTZ	4.10
5	TPSSTPSS/aug-cc-pVTZ	4.73
1	TPSSKCIS/6-31+G*	4.76
2	TPSSTPSS/6-31+G*	4.77
3	B3PW91/6-31G*	4.79
Hydrogen Bond Interaction Energy		(kcal/mol)
1	MPWLYP/aug-cc-pVTZ	0.31
2	B1LYP/6-31++G*	0.33
3	MPWLYP/aug-cc-pVDZ	0.33
4	B1LYP/6-31+G*	0.34
5	PBE1KCIS/aug-cc-pVTZ	0.36
1	B1LYP/6-31++G*	0.33
2	B1LYP/6-31+G*	0.34
3	B3LYP/6-31++G*	0.36
Conformational Energy		(% error)
1	MPWB95/cc-pVTZ	7.90
2	B1B95/aug-cc-pVDZ	8.10
3	BB1K/aug-cc-pVDZ	8.30
4	PBEP86/cc-pVTZ	8.30
5	BB95/cc-pVTZ	8.60
1	B1B95/6-31++G*	12.20
2	MPWB95/6-31++G*	12.40
3	B1B95/6-31+G*	12.50

SRBH		(kcal/mol)
1	BB1K/aug-cc-pVTZ	1.05
2	BB1K/cc-pVTZ	1.31
3	BB1K/aug-cc-pVDZ	1.69
4	BB1K/6-31+G*	1.95
5	BB1K/6-31++G*	2.58
1	BB1K/6-31+G*	1.95
2	BB1K/6-31++G*	2.58
3	BB1K/6-31G*	2.60
LSBH		(kcal/mol)
1	B1LYP/aug-cc-pVTZ	2.575
2	B1LYP/cc-pVTZ	2.591
3	B1LYP/6-31++G*	2.631
4	B1LYP/6-31+G*	2.637
5	B3LYP/aug-cc-pVTZ	3.102
1	B1LYP/6-31++G*	2.631
2	B1LYP/6-31+G*	2.637
3	B1LYP/6-31G*	3.123

physical property considered here, with the exception of conformational energies, the best results obtained with Pople-type basis sets are comparable to the best results produced by the larger Dunning-type bases.

One of the main goals of this survey is to get a rough estimate of a functional's performance in terms of its ability to describe all of the properties considered in this study. In order to accomplish this goal we compare the average functional ranks and standard deviations for each of the functionals studied in this work. The average functional rank is given as the mean of a functional's rank for all of the properties considered here and the standard deviation was also calculated.

Table 6 lists the average functional ranks and standard deviations of the fifteen functionals with the lowest average ranks for the 6-31+G* and aug-cc-pVDZ basis sets. For both basis sets there are five hybrid-meta-GGA and three meta-GGA functionals

represented in the “top fifteen”. The top fifteen of the 6-31+G* basis also included four hybrid-GGA and three GGA functionals, while the top performers from the aug-cc-pVDZ set included five hybrid-GGA and two GGA functionals. In the aug-cc-pVDZ group, each of the “top five” functionals in terms of average functional rank contains “exact exchange” terms. Whereas the only three of the “top five” of the 6-31+G* set contain an “exact-exchange” term. Also, for both basis sets, the only GGA functional to rank in the top ten is MPWPW91/aug-cc-PVDZ.

Table 5-23 lists the fifteen best functionals for the 6-31+G* basis set along with their unsigned errors for each of the properties considered in this work, for purposes of comparison, the lowest and highest unsigned errors for each property are given, as well as the mean unsigned error averaged over all of the functionals in this study. For the 6-31+G* basis set, the B1B95 functional obtains the lowest average functional rank with a value of 10.7. However, the standard deviation of this functional is fairly high with a value of 11.9, since the method performs very well for some properties and relatively poorly for others, as can be seen in Table 5-24. Other functionals that perform notably well are B98, TPSSTPSS, TPSS1KCIS, and PBE1PBE; each of these functionals gives reasonably good results for all of the physical properties here (with the possible exception of vibrational frequencies).

Table 5-25 lists the fifteen best functionals for the aug-cc-pVDZ basis set along with their unsigned errors for each of the properties considered in this work in the same manner as was done for the 6-31+G* basis. For this basis set there are a number of functionals that perform very well in terms of giving a good description of each of the physical properties in this work. The B98 functional has the lowest average functional

ranking with a value of 10.1 (standard deviation = 8.8). When paired with aug-cc-pVDZ, B98 ranks in the top eleven functionals for all properties except HOF and vibrational frequency. B98's predicted heat of formation is in error by an average of 18.38 kcal/mol. TPSS1KCIS, which ranks as third best with the DZ basis, predicts HOF very well, but is less accurate for electron affinity, conformational energy, and vibrational frequency. Other functionals of note are B1B95, PBE1PBE, and B3LYP.

Table 5-23. Average functional rankings and standard deviations for the top fifteen functionals along with 6-31+G* and aug-cc-pVDZ basis sets.

	6-31+G*	Avg. Rank	Std. Dev.
1	B1B95	10.7	11.9
2	B98	11.9	7.5
3	TPSSKCIS	13.6	8.4
4	TPSSTPSS	13.7	8.4
5	PBE1PBE	13.8	10.9
6	B3LYP	13.9	9.0
6	MPWB95	13.9	11.2
8	TPSS1KCIS	14.0	8.2
9	B3PW91	14.2	9.2
9	BB1K	14.2	12.4
11	MPW1KCIS	14.8	10.5
12	MPWPW91	15.8	5.6
13	PBEPW91	16.6	8.5
14	PBE1KCIS	16.7	9.2
15	MPWPBE	16.8	5.3
	aug-cc-pVDZ		
1	B98	10.1	8.8
2	B1B95	11.7	12.2
3	TPSS1KCIS	12.0	8.1
4	PBE1PBE	12.2	10.0
5	B3LYP	12.3	9.2
6	PBE1KCIS	12.8	10.9
7	TPSSTPSS	13.3	6.3
8	TPSSKCIS	13.6	7.6
9	B3PW91	13.8	9.0
10	MPWPW91	15.3	5.9
11	MPWPBE	15.7	6.7
12	MPW1KCIS	15.9	9.9
13	BB95	16.2	9.3
13	B1LYP	16.2	13.1
15	BB1K	16.7	13.8

Table 5-24. Performances of the fifteen highest-ranking functionals paired with the 6-31G* basis.

	HOF	IP	EA	H-bond	Freq	Length	Angle	Conf E	SRBH	LSBH
B1B95	9.94	4.81	5.07	0.64	104.1	0.0074	1.23	12.47	2.64	4.16
B98	13.47	5.05	3.83	0.43	88.6	0.0094	1.29	14.59	3.81	4.28
TPSSKCIS	4.76	5.99	4.47	0.43	65.6	0.0135	1.26	14.07	6.41	7.04
TPSSTPSS	4.77	5.52	4.83	0.47	65.8	0.0135	1.23	13.73	7.33	7.25
PBE1PBE	5.94	5.34	5.15	0.77	103.4	0.0079	1.23	15.45	3.92	5.49
B3LYP	14.03	5.29	3.91	0.38	84.4	0.0093	1.29	16.54	4.36	3.32
MPWB95	18.31	4.53	3.12	0.44	48.0	0.0161	1.31	12.66	8.57	8.41
TPSS1KCIS	12.06	5.30	5.07	0.43	79.8	0.0090	1.25	16.23	4.37	5.12
B3PW91	8.32	5.48	4.39	0.56	92.5	0.0081	1.26	16.46	3.73	5.01
BB1K	17.34	5.33	6.40	0.40	138.2	0.0098	1.28	12.79	1.95	3.38
MPW1KCIS	7.28	4.86	3.98	0.40	75.9	0.0101	1.60	16.71	5.71	5.41
MPWPW91	8.57	5.16	3.80	0.46	49.7	0.0157	1.29	15.67	7.78	8.17
PBEPW91	17.18	5.10	3.67	0.89	48.8	0.0163	1.28	14.84	8.62	8.41
PBE1KCIS	15.20	4.95	4.22	0.63	91.0	0.0086	1.57	16.18	5.09	5.07
MPWPBE	8.76	5.16	3.87	0.47	48.9	0.0159	1.29	15.79	7.83	8.32
Lowest Err.	TPSSKCIS	MPWB95	MPWB95	B1LYP	G96P86	B1B95	cSVWN5	B1B95	BB1K	B1LYP
Value	4.76	4.53	3.12	0.34	49	0.007	1.28	12.00	1.95	2.64
Highest Err.	SPL	cSVWN5	cSVWN5	SPL	BB1K	cSVWN5	VSXC	VSXC	SPL	SPL
Value	133.7	19.08	14.8	6.21	142	0.025	1.56	44	16.75	12.04
Avg. Err.	15.82	5.78	4.54	0.93	66.01	0.014	1.33	16.24	6.74	6.62

Notes: Errors given in the following units: bond length (Å), bond angle (degrees), frequency (cm⁻¹), ionization potential (kcal/mol), electron affinity (kcal/mol), heat of formation (kcal/mol), hydrogen-bond interaction energy (kcal/mol), conformational energy (percent error), reaction barrier height (kcal/mol). Average errors include all 37 density functionals considered in this work.

Table 5-25. Performances of the fifteen highest ranked density functional methods paired with the aug-cc-pVDZ basis set.

	HOF	IP	EA	H-bond	Freq	Length	Angle	Conf E	SRBH	LSBH
B98	18.38	4.90	3.42	0.40	73.9	0.0114	1.24	9.85	4.98	4.68
B1B95	14.13	4.53	4.54	1.13	88.3	0.0093	1.20	8.06	3.72	4.67
TPSS1KCIS	8.35	5.30	4.64	0.55	68.5	0.0109	1.21	11.70	5.48	5.58
PBE1PBE	8.82	5.27	4.66	0.56	86.5	0.0104	1.17	10.71	5.17	5.92
B3LYP	18.66	5.28	3.78	0.63	70.5	0.0111	1.22	12.34	5.37	4.12
PBE1KCIS	11.63	4.84	3.79	0.37	82.2	0.0109	1.54	11.84	6.37	4.71
TPSSTPSS	8.72	5.53	4.49	0.57	50.9	0.0159	1.23	8.99	8.61	8.02
TPSSKCIS	8.31	5.96	4.13	0.55	52.1	0.0157	1.25	9.58	7.51	7.64
B3PW91	12.02	5.42	4.08	1.09	77.6	0.0104	1.15	11.80	4.96	5.52
MPWPW91	8.58	5.36	4.03	0.69	51.1	0.0177	1.27	11.17	8.85	8.77
MPWPBE	8.91	5.27	3.89	0.72	50.9	0.0180	1.27	11.29	8.92	8.85
MPW1KCIS	12.57	4.78	3.58	0.76	70.9	0.0121	1.58	12.64	7.00	5.92
BB95	9.94	5.00	3.85	1.58	52.6	0.0179	1.31	8.98	8.62	8.50
B1LYP	34.07	5.93	4.63	0.68	77.6	0.0092	1.20	12.61	4.15	3.19
BB1K	21.29	6.50	5.94	0.93	119.3	0.0107	1.23	8.31	1.69	4.54
Lowest Err.	PW91LYP	B1B95	B98	PBEKCIS	G96LYP	B1LYP	cSVWN5	B1B95	BB1K	B1LYP
Value	8.24	4.53	3.43	0.32	48	0.009	1.25	8.1	1.69	3.19
Highest Err.	SPL	cSVWN5	B3P86	cSVWN5	BB1K	BLYP	BLYP	VSXC	SPL	SPL
Value	128.62	19.06	13.99	10.26	119	0.029	1.6	51.5	18.2	12.54
Avg. Error	16.23	5.77	4.77	1.00	62.2	0.016	1.30	12.33	7.84	7.15

Notes: Errors given in the following units: bond length (Å), bond angle (degrees), frequency (cm⁻¹), ionization potential (kcal/mol), electron affinity (kcal/mol), heat of formation (kcal/mol), hydrogen-bond interaction energy (kcal/mol), conformational energy (percent error), reaction barrier height (kcal/mol). Average errors include all 37 density functionals considered in this work.

CHAPTER 6 CONCLUSIONS

This chapter provides a rough overview of the work presented in this dissertation. Since the three studies presented in this work are for the most part unrelated, the individual chapters contain more in-depth discussion. These research projects demonstrate the usefulness of computational methods to address important biological and chemical issues in the human body. The studies performed on HAH1 and AAP are only possible after much information has been obtained about these systems through physical experimentation. With this in mind, it is a combination of computational and analytical methods that provide a complete description of the geometric and energetic phenomena that occur in nature.

Chapter two outlined the techniques used in the study of metalloenzymes. These methods included density functional and *ab initio* quantum mechanics calculations on model systems, molecular mechanics, and molecular dynamics simulations of large-scale solvated protein systems. The results of these studies are shown in chapters 3 and 4.

HAH1 is a key protein for Cu homeostasis in the human body. Improper function of HAH1 or of its target protein the Wilson's disease ATPase results in accumulation of too much Cu or insufficient amount of Cu in the cell, which both cause death. One of the key unanswered questions involving Cu(I) homeostasis is the mechanism by which the metal ion is transferred from the donor protein to the target active site. The study described here provides evidence that Cys 17 of the MNK4 target domain is energetically favored to be the first target residue to bind Cu(I).

The AAP system has also been associated with several serious illnesses. One subject of investigation in this work was the di-Zinc bridging species in the AAP active site. The electronic structure of the active site was fully characterized by QM calculations, and the contribution of the 1st shell residues to Zn ion coordination was determined. The results from this work suggest that the bridging species is most like a OH⁻ group which then becomes a terminal group upon substrate binding.

The fifth chapter details the large-scale survey of the accuracy of well-known density functional methods for nine molecular and intermolecular properties. Comprising more than 150,000 individual calculations, this study is believed to be the most comprehensive of its kind ever performed. The newer hybrid-meta-GGA methods showed the most promise, with the MPW1KCIS density functional receiving the best overall score for accuracy over all of the properties that were surveyed.

Although many of the methods employed in these studies are approximations, they have all been well validated and can be used to obtain highly reliable and accurate information on the conformations and energetics of chemical systems. The simulation of metalloenzymes is an example of the kind of computational work that has become possible over the last decade thanks to advances in analytical and computational instrumentation, and soon these simulations may be fully quantum mechanical in nature. The continued development of parameters for metal ions and metal-binding active sites within proteins will allow for further identification of the mechanisms of metal-catalyzed processes in the body.

APPENDIX
Gaussian Keywords

opt	Perform a geometry optimization
freq=noraman	Perform a single-point frequency calculation for vibrational frequencies and ground state thermodynamical properties. Noraman is selected to skip the calculation of Raman spectral data.
IOP(7/33=1)	Has bonding force constants included in the Gaussian output file
pop(mk,readradii)	Perform a population analysis calculation for atomic charges using the Merz-Kollman-Singh method. Readradii allows the user to define ionic radii that are not listed in the Gaussian library of ionic radii. This was used to define the ionic radius of Cu(I).
IOP(6/33=2)	Used in Gaussian 98 to format the output from charge calculations into AMBER-readable format
Gen 6D 7F	Specifies the use of multiple basis sets for different atoms within a system. 6D 7F specifies the use of 6 Cartesian d-orbitals and 7 Cartesian f-orbitals instead.
scrf(pcm,solvent=water)	For self-consistent reaction field calculations using the polarizable continuum model for implicit solvation using water as the solvent,
int(grid=ultrafine)	Requests a pruned (99,590) grid that is finer than the default (75,302) grid
scf(xqc,maxcycle=N)	xqc requests the quadratically convergent SCF procedure for energy convergence with an extra qc step in case the first-order scf did not converge. This is computationally more expensive than the default method. maxcycle sets the maximum number of iterations allowed before the program exits due to convergence failure
ts	Requests geometry optimization to a transition state rather than to a local minimum state.
calcall	Requests the calculation of force constants at every point along the optimization pathway. This also increases the computational expense of the experiment.
noeigentest	For use in TS searches. This skips the portion of the calculation where Gaussian checks to ensure the existence of only one negative eigenvalue for the structure being optimized to a TS.

LeAP Commands in AMBER

loadpdb	Load PDB structure into LeAP and protonate the structure
loadamberparams	Load AMBER parameter files for the system
edit	Edit the unit that has been created in LeAP. Charges, atom types, atom names, and perturbation flags can be edited in this manner.
check	Check the unit for errors
saveoff	Save all atom names, atom types, charges, and perturbation information into a library for a unit.
loadoff	Load information about a unit that was previously created and stored.
solvateBox	Add a solvent box around a unit
saveamberparm	Save the any non-default parameters that have been imposed on the system by the previously loaded AMBER parameter files.
saveamberparmpert	Save perturbation information and the parameters associated with the perturbation.

LIST OF REFERENCES

- (1) Wernimont, A. K.; Huffman, D. L.; Lamb, A. L.; O'Halloran, T. V.; Rosenzweig, A. C. *Nat Struct Biol* **2000**, *7*, 766-771.
- (2) Anastassopoulou, A.; Banci, L.; Bertini, I.; Cantini, F.; Katsari, E.; Rosato, A. *Biochem* **2004**, *43*, 13046-13053.
- (3) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J Am Chem Soc* **1985**, *107*, 3902-3909.
- (4) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- (5) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 43.
- (6) Case, D. A.; Darden, T. A.; Cheatham, T. E. I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M. J.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A. E.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A.; 9 ed.; University of California, San Francisco: San Francisco, 2006.
- (7) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Chesseman, J. R.; Zakrzewski, V. G.; Montgomery Jr., J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; AlLoham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A.; C.01 ed.; Gaussian Inc.: Pittsburgh, PA, 2003.
- (8) Ni, B.; Kramer, J. R.; Werstiuk, N. H. *J. Phys. Chem. A* **2003**, *107*, 8949-8954.
- (9) Olsson, M. H. M.; Ryde, U. *J Am Chem Soc* **2001**, *123*, 7866-7876.
- (10) Ryde, U.; Olsson, M. H. M.; Roos, B. O.; Borin, A. C. *Theo. Chem. Acct.* **2001**, *105*, 452-462.
- (11) Siegbahn, P. E. M.; Blomberg, M. R. A. *Ann. Rev. Phys. Chem* **1999**, *50*, 221-249.

- (12) Rulisek, L.; Solomon, E. I.; Ryde, U. *Inorg Chem* **2005**, *44*, 5612-5628.
- (13) Schäfer, A.; Horn, H.; Ahlrichs, R. *J Chem Phys* **1992**, *97*.
- (14) Mitlin, A. V.; Baker, J.; Pulay, P. *J Chem Phys* **2003**, *118*, 7775-7782.
- (15) Rassolov, V. A.; Pople, J. A.; Ratner, J. A.; Windus, T. L. *J Chem Phys* **1998**, *109*, 1223-1229.
- (16) Molecular Simulations, I.; 3.7 ed., 2000.
- (17) Besler, B. H.; Merz, K. M. J.; Kollman, P. A. *J Comp Chem* **1990**, *11*, 431.
- (18) Singh, U. C.; Kollman, P. A. *J Comp Chem* **1984**, *5*, 129.
- (19) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J Am Chem Soc* **1995**, *117*, 5179-5197.
- (20) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E. I.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R.; Duan, Y.; Pitera, J.; Massova, I.; Kollman, P. A. *AMBER 7 Users' Manual*; University of California: San Francisco, 2002.
- (21) Wollacott, A. M. In *Chemistry*; Pennsylvania State University: State College, PA, 2005.
- (22) Rossi, K. In *Chemistry*; Pennsylvania State University: State College, PA, 1994.
- (23) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; 1st ed.; John Wiley & Sons, Ltd.: Chichester, England, 2002.
- (24) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J Comp Chem* **1992**, *13*, 1011.
- (25) Roux, B. *Comp Phys Conn* **1995**, *91*, 275-282.
- (26) Chodera, J. D.; Swope, W. C.; Pitera, J.; Seok, C.; Dill, K. *JCTC* **2005**, *Submitted*.
- (27) Harrison, M. D.; Jones, C. E.; Solioz, M.; Dameron, C. T. *TIBS* **2000**, *25*, 29-32.

- (28) Rae, T. D.; Schmidt, P. J.; Pufahl, R. A.; Culotta, V. C.; O'Halloran, T. V. *Science* **1999**, *284*, 805-808.
- (29) O'Halloran, T. V.; Culotta, V. C. *J Biol Chem* **2000**, *275*, 25057-25060.
- (30) Gaggelli, E.; Kozlowski, H.; Valensin, D.; Valensin, G. *Chem. Rev.* **2006**, *106*, 1995-2044.
- (31) Huffman, D. L.; O'Halloran, T. V. *Ann. Rev. Biochem.* **2001**, *70*, 677-701.
- (32) Lieberman, R. L.; Rosenzweig, A. C. In *Comprehensive Coordination Chemistry II: from biology to nanotechnology*; 1st ed., 2004, pp 195-211.
- (33) Puig, S.; Thiele, D. J. *Curr Opin Chem Biol* **2002**, *6*, 171-180.
- (34) Hamza, I.; Schaefer, M.; Klomp, L. W. J.; Gitlin, J. D. *Proc. Natl. Acad. Sci.* **1999**, *96*, 13363-13368.
- (35) Mercer, J. F. B.; Barnes, N.; Stevenson, J.; Strausak, D.; Llanos, R. M. *BioMetals* **2003**, *16*, 175-184.
- (36) Valentine, J. S.; Gralla, E. B. *Science* **1997**, *278*, 817-818.
- (37) Rosenzweig, A. C. *Acc. Chem. Res.* **2001**, *34*, 119-128.
- (38) Gresh, N.; Policar, C.; Giessner-Prettre, C. *J. Phys. Chem. A* **2002**, *106*, 5660-5670.
- (39) Lamb, A. L.; Torres, A. S.; O'Halloran, T. V.; Rosenzweig, A. C. *Biochem* **2000**, *39*, 14720-14727.
- (40) Portnoy, M. E.; Rosenzweig, A. C.; Rae, T.; Huffman, D. L.; O'Halloran, T. V.; Culotta, V. C. *J. Biol. Chem.* **1999**, *274*, 15041-15045.
- (41) Gitschier, J.; Moffat, B.; Reilly, D.; Wood, W. I.; Fairbrother, W. J. *Nat Struct Biol* **1998**, *5*, 47-54.
- (42) Lamb, A. L.; Torres, A. S.; O'Halloran, T. V.; Rosenzweig, A. C. *Nat Struct Biol* **2001**, *8*, 751-755.
- (43) Zerner, M. C.; Bacon, A. D. *Theo. Chem. Acct.* **1979**, *53*, 21-54.
- (44) Bertini, I.; Gray, H.; Lippard, S.; Valentine, J. S. *Bioinorganic Chemistry*; University Science Books: Sausalito, CA, 1994.
- (45) O'Halloran, T. V. *Science* **1993**, *261*, 715-725.

- (46) Radisky, D.; Kaplan, J. *J. Biol. Chem.* **1999**, *274*, 4481-4484.
- (47) Rosenzweig, A. C.; Huffman, D. L.; Hou, M. Y.; Wernimont, A. K.; Pufahl, R. A.; O'Halloran, T. V. *Structure* **1999**, *7*, 605-617.
- (48) Larin, D.; Mekios, C.; Das, K.; Ross, B.; Yang, A.; Gilliam, T. C. *J Biol Chem* **1999**, *274*, 28497-28504.
- (49) Lamb, A. L.; Pufahl, R. A.; O'Halloran, T. V.; Culotta, V. C.; Rosenzweig, A. C. *Nat Struct Biol* **1999**, *6*, 724-729.
- (50) Lamb, A. L.; Wernimont, A. K.; Pufahl, R. A.; O'Halloran, T. V.; Rosenzweig, A. C. *Biochem* **2000**, *39*, 1589-1595.
- (51) Ochterski, J. W.; Gaussian, Inc., 2000.
- (52) Pufahl, R. A.; Singer, C. P.; Peariso, K. L.; Lin, S.-J.; Schmidt, P. J.; Fahrni, C. J.; Culotta, V. C.; Penner-Hahn, J. E.; O'Halloran, T. V. *Science* **1997**, *278*, 853-856.
- (53) M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, Ö. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, P. Salvador, J. J. Dannenberg, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komáromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople; A.11 ed.; Gaussian, Inc.: Pittsburgh, PA, 1998.
- (54) Arnesano, F.; Banci, L.; Bertini, I.; Bonvin, A. M. *Structure* **2004**, *12*, 669-676.
- (55) Comba, P.; Remenyi, R. *J Comp Chem* **2002**, *23*, 697-705.
- (56) Christianson, D. W.; Lipscomb, W. N. *Journal of the American Chemical Society* **1980**, *108*, 4998-5003.
- (57) Holz, R. C. *Coordination Chemistry Reviews* **2002**, *232*, 5-26.
- (58) Christianson, D. W.; Lipscomb, W. N. *Accounts of Chemical Research* **1989**, *22*, 62-69.

- (59) Desmarais, W. T.; Bienvenue, D. L.; Bzymek, K. P.; Holz, R. C.; Petsko, G. A.; Ringe, D. *Structure* **2002**, *10*, 1063-1072.
- (60) Matthews, B. W. *Accounts of Chemical Research* **1988**, *21*.
- (61) Lowther, W. T.; Matthews, B. W. *Chemical Reviews* **2002**, *102*, 4581-4607.
- (62) Weston, J. *Chemical Reviews* **2005**, ASAP Article.
- (63) Prescott, J. M.; Wilkes, S. H. *Methods of Enzymology* **1976**, *45*, 530-543.
- (64) Prescott, J. M.; Wagner, F. W.; Holmquist, B.; Vallee, B. L. *Biochemistry* **1985**, *24*, 5350-5356.
- (65) Bennett, B.; Antholine, W. E.; D'souza, V. M.; Chen, G. J.; Ustinyuk, L.; Holz, R. C. *Journal of the American Chemical Society* **2002**, *124*, 13025-13034.
- (66) Stamper, C.; Bennett, B.; Edwards, T.; Holz, R. C.; Ringe, D.; Petsko, G. *Biochemistry* **2001**, *40*, 7035-7046.
- (67) Chevrier, B.; Schalk, C.; D'Orchymont, H.; Rondeau, J.-M.; Moras, D.; Tarnus, C. *Structure* **1994**, *2*, 283-291.
- (68) Bertini, I.; Luchinat, C.; Rosi, M.; Sgamellotti, A.; Tarantelli, F. *Inorg Chem* **1990**, *29*, 1460-1463.
- (69) Chevrier, B.; D'Orchymont, H.; Schalk, C.; Tarnus, C.; Moras, D. *Eur J Biochem* **1996**, *237*, 393-398.
- (70) De Paola, C. C.; Bennett, B.; Holz, R. C.; Ringe, D.; Petsko, G. A. *Biochemistry* **1999**, *38*, 9048-9053.
- (71) Schürer, G.; Lanig, H.; Clark, T. *Biochemistry* **2004**, *43*, 5414-5427.
- (72) Baker, J. O.; Prescott, J. M. *Biochemistry* **1983**, *22*, 5322-5331.
- (73) Bienvenue, D. L.; Bennett, B.; Holz, R. C. *J Inorg Biochem* **2000**, *78*, 43-54.
- (74) Chan, W. W.; Dennis, P.; Demmer, W.; Brand, K. *Journal of Biological Chemistry* **1982**, *257*, 7955-79757.
- (75) Wilkes, S. H.; Prescott, J. M. *Journal of Biological Chemistry* **1983**, *258*, 13517-13521.

- (76) Wilkes, S. H.; Prescott, J. M. *Journal of Biological Chemistry* **1987**, *262*, 8621-8625.
- (77) Chen, G.; Edwards, T.; D'souza, V. M.; Holz, R. C. *Biochemistry* **1997**, *36*, 4278-4286.
- (78) Martell, A. E.; Hancock, R. D.; Plenum Press: New York, 1996, p 199.
- (79) Christianson, D. W.; Cox, J. D. *Annu Rev Biochem* **1999**, *68*, 33-57.
- (80) Bennett, B.; Holz, R. C. *Journal of the American Chemical Society* **1997**, *119*, 1923-1933.
- (81) Bennett, B.; Holz, R. C. *Biochemistry* **1997**, *36*, 9837-9846.
- (82) Bennett, B.; Holz, R. C. *Journal of the American Chemical Society* **1998**, *120*, 12139-12140.
- (83) Christianson, D. W.; Alexander, R. S. *Journal of the American Chemical Society* **1989**, *111*, 6412-6419.
- (84) Stamper, C. C.; Bienvenue, D. L.; Bennett, B.; Ringe, D.; Petsko, G. A.; Holz, R. C. *Biochemistry* **2004**, *43*, 9620-9628.
- (85) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; J. A. Montgomery, J.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; A.1 ed.; Gaussian, Inc: Pittsburgh, PA, 2003.
- (86) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; 1st, revised ed.; Dover Publications, Inc.: Mineola, NY, 1989.
- (87) Levine, I. N. *Quantum Chemistry*; 5th ed.; Prentice Hall: Upper Saddle River, NJ, 2000.

- (88) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989; Vol. 16.
- (89) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864-B871.
- (90) Kohn, W. In *Highlights of Condensed-Matter Theory*; Bassani, F., Fumi, F., Tosi, M. P., Eds.; North-Holland: Amsterdam, 1985.
- (91) Scuseria, G. E.; Staroverov, V. N. In *Theory and Applications of Computational Chemistry: The First 40 Years (A Volume of Technical and Historical Perspectives)*; C. E. Dykstra, G. F., K. S. Kim, G. E. Scuseria, Ed.; Elsevier: Amsterdam, 2005, pp 669-724.
- (92) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133-A1138.
- (93) Roothan *Rev. Mod. Phys.* **1951**, *23*, 69.
- (94) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
- (95) Adamo, C.; Barone, V. J. *Chem. Phys. Lett.* **1997**, *274*, 242-250.
- (96) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098-3100.
- (97) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785-789.
- (98) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623-11627.
- (99) Hertwig, R. H.; Koch, W. *Chem. Phys. Lett.* **1997**, *268*, 345-351.
- (100) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865-3868.
- (101) Slater, J. C. *Quantum Theory of Molecular and Solids. Vol. 4: The Self-Consistent Field for Molecular and Solids*; McGraw-Hill: New York, 1974.
- (102) Vosko, S. H.; Wilk, L.; Nusair, M. *Canadian Journal of Physics* **1980**, *58*, 1200-1211.
- (103) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822-8824.
- (104) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244-13249.
- (105) Perdew, J. P.; Chevary, J. a.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Physical Review B* **1992**, *46*, 6671-6687.

- (106) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- (107) Riley, K. E.; Brothers, E. N.; Ayers, K. B.; Merz, K. M. *JCTC* **2005**, *1*, 546-553.
- (108) Schmider, H. L.; Becke, A. D. *J. Chem. Phys.* **1998**, *108*, 9624-9631.
- (109) Voorhis, T. V.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400-410.
- (110) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040-1046.
- (111) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664-675.
- (112) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129-12137.
- (113) Rey, J.; Savin, A. *Int. J. Quantum Chem.* **1998**, *69*, 581-590.
- (114) Krieger, J. B.; Chen, J.; Iafrate, G. J.; Savin, A. In *Electron Correlations and Materials Properties*; Gonis, A., Kioussis, N., Eds.; Plenum: New York, 1999, p 463.
- (115) Toulouse, J.; Savin, A.; Adamo, C. *J. Chem. Phys.* **2002**, *117*, 10465-10473.
- (116) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2715-2719.
- (117) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 43-52.
- (118) Zhao, Y.; Truhlar, D. G. *JCTC* **2005**, *1*, 415.
- (119) Gill, P. M. W. *Mol. Phys.* **1996**, *89*, 433-445.
- (120) Zhao, Y.; González-García, N.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 2012-2018.
- (121) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 6264-6271.
- (122) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939-947.
- (123) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007-1023.

- (124) Gordon, M. S.; Binkley, J. S.; Pople, J. A.; Pietro, W. J.; Hehre, W. J. *J. Am. Chem. Soc.* **1982**, *104*, 2797-2803.
- (125) Hehre, W. J.; Ditchfie, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257-2261.
- (126) Pietro, W. J.; Francl, M. M.; Hehre, W. J.; Defrees, D. J.; Pople, J. A.; Binkley, J. S. *J. Am. Chem. Soc.* **1982**, *104*, 5039-5048.
- (127) Hinchliffe, A. *Modeling Molecular Structures*; 2nd ed.; Wiley & Sons, Ltd.: Chichester, 2000.
- (128) Foresman, J. B.; Frisch, A. *Exploring Chemistry with Electronic Structure Methods*; 2nd ed.; Gaussian, Inc.: Pittsburgh, PA, 1996.
- (129) Frisch, A.; Frisch, M. J.; Trucks, G. W. *Gaussian 03 User's Reference*; Gaussian, Inc.: Pittsburgh, PA, 2003.
- (130) Bauschlicher, C. W. *Chem. Phys. Lett.* **1995**, *246*, 40-44.
- (131) Johnson, B. G.; Gill, P. M. W.; Pople, J. A. *J. Chem. Phys.* **1992**, *97*, 7846-7848.
- (132) Johnson, B. G.; Gill, P. M. W.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 5612-5626.
- (133) Neugebauer, A.; Häfelfinger, G. *Journal of Molecular Structure-Theochem* **2002**, *578*, 229.
- (134) Neugebauer, A.; Häfelfinger, G. *Journal of Molecular Structure-Theochem* **2002**, *585*, 35-47.
- (135) Raymond, K. S.; Wheeler, R. a. *J. Comp. Chem.* **1999**, *20*, 207-216.
- (136) Scheiner, A. C.; Baker, J.; Andzelm, J. W. *J. Comp. Chem.* **1997**, *18*, 775-795.
- (137) Wang, N. X.; Wilson, A. K. *J. Chem. Phys.* **2004**, *121*, 7632-7646.
- (138) Wang, N. X.; Wilson, A. K. *Molecular Physics* **2005**, *103*, 345-358.
- (139) Schaftenaar, G.; Noordik, J. H. *J. Comput.-Aided Mol. Design* **2000**, *14*, 123-124.
- (140) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063-1079.

- (141) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 42-55.
- (142) Brothers, E. N.; Merz, K. M. *J. Phys. Chem. A* **2004**, *108*, 2904-2911.
- (143) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374-7383.
- (144) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2005**, *123*, 124107/124101-124107/124112.
- (145) Curtiss, L. A.; Redfern, P. C.; Rossolov, V.; Kedziora, Z.; Pople, J. A. *J. Chem. Phys.* **2001**, *114*, 9287-9295.
- (146) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029-5036.
- (147) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 8996-8999.
- (148) Mole, S. J.; Zhou, X.; Liu, R. *J. Phys. Chem.* **1996**, *100*, 14665-14671.
- (149) Rabuck, A. D.; Scuseria, G. E. *Chem. Phys. Lett.* **1999**, *309*, 450-456.
- (150) Brothers, E. N.; Scuseria, G. E. *J. Chem. Theo. and Computation* **2006**, *2*, 1045-1049.
- (151) Riley, K. E.; Op't Holt, B. T.; Merz, K. M. *J. Chem. Theo. and Computation* **2006**, *In Press*.
- (152) Tsuzuki, S.; Lüthi, H. P. *J. Chem. Phys.* **2001**, *114*, 3949-3957.
- (153) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comp. Chem.* **2002**, *23*, 1601-1622.
- (154) Helkier, A.; Klopper, W.; Helgaker, T.; Jorgensen, P.; Taylor, P. R. *J. Chem. Phys.* **1999**, *111*, 9157.
- (155) Hobza, P.; Sponer, J.; Reschel, T. *J. Comp. Chem.* **1995**, *16*, 1315-1325.
- (156) Ireta, J.; Neugebauer, J.; Scheffler, M. *J. Phys. Chem. A* **2004**, *108*, 5692-5698.
- (157) Paizs, B.; Suhai, S. *J. Comp. Chem.* **1998**, *19*, 575-584.
- (158) Rabuck, A. D.; Scuseria, G. E. *Ther. Chem. Acc.* **2000**, *104*, 439-444.

- (159) Rappe, A. K.; Bernstein, E. R. *J. Phys. Chem. A* **2000**, *104*, 6117-6128.
- (160) Tuma, C.; Boses, A. D.; Handy, N. C. *Phys. Chem. Chem. Phys.* **1999**, *1*, 3939-3947.
- (161) Xu, X.; Goddard, W. A. *J. Chem. Phys. A* **2004**, *108*, 2305-2313.
- (162) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384-1388.
- (163) Stamant, A.; Cornell, W. D.; Kollman, P. A.; Holgren, T. A. *J. Comput. Chem.* **1995**, *16*, 1483.
- (164) Dickson, R. M.; Becke, A. D. *J. Chem. Phys.* **2005**, *123*, Art. No. 111101/111101-111101/111103.
- (165) Dybala-Defratyka, A.; Paneth, P.; Pu, J. Z.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2475-2486.
- (166) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811.
- (167) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2002**, *106*, 842-846.
- (168) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553-566.
- (169) Freile, M. L.; Risso, S.; Curaqueo, A.; Zamora, M. A.; Enriz, R. D. *J. Molec. Struct.-Theochem.* **2005**, *2005*, 107.
- (170) Halgren, T. *J. Comp. Chem.* **1999**, *20*.
- (171) Zhong, H. Z.; Stewart, E. L.; Kontoyianni, M.; Bowen, J. P. *J. Chem. Theo. and Computation* **2005**, *1*, 230.

BIOGRAPHICAL SKETCH

I was born in Chicago in 1979 when my parents worked at the hospital and while my dad was a graduate student at Illinois. I attended three high schools, the last of which, the Alabama School of Mathematics and Science, provided a strong foundation in chemistry and math that carried me into college. I spent four years at the University of South Alabama obtaining a BS in chemistry and working in the laboratory of Doctor Andrej Wierzbicki. He was the first to suggest that I continue my education in chemistry by attending graduate school. I enrolled at Pennsylvania State University in 2001, where I joined Doctor Kennie Merz and his group with the intention of using quantum chemistry to address complex problems that could not be solved in a wet chemistry lab. After graduating from the University of Florida, I joined Doctor Ed Solomon at Stanford University as a postdoctoral researcher in his group.

I married Jenny in 2004, and she has been wonderfully supportive of me. I would have to say that the toughest part of graduate school for me was getting out of bed every morning and trying to make a little bit of progress each day. Graduate school guarantees no instant gratification, but offers a lifetime of rewards upon successful completion.