RIJKSUNIVERSITEIT GRONINGEN

DENSITY FUNCTIONAL THEORY APPLIED TO COPPER PROTEINS

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
dinsdag 10 september 2002
om 13.15 uur

door

Marcel Swart

geboren op 1 maart 1971
te Zutphen

Promotores:

prof. dr. J.G. Snijders
prof. dr. H.J.C. Berendsen
prof. dr. G.W. Canters

Beoordelingscommissie:

prof. dr. B.W. Dijkstra
prof. dr. K. Lammertsma
prof. dr. A.E. Mark

voor Elvira

in lieve herinnering aan papa en mama

Dit proefschrift werd gedrukt door drukkerij Febodruk, Enschede.
Oplage: 465 exemplaren, waarvan vijftien in hardcover gebonden.

Het wapen op de voorkant van dit proefschrift is het familiewapen van de familie (de) Swart.

Het promotieonderzoek beschreven in dit proefschrift werd financieel mogelijk gemaakt door de stichting Scheikundig Onderzoek Nederland (SON, tegenwoordig NWO/CW) en Unilever Research Vlaardingen.

Preface

In November 1996, the Netherlands Foundation for Chemical Research (SON, now NWO/CW) and Unilever decided to initiate a joint research program called *Computational chemistry of biosystems*, as this field was considered scientifically challenging and promising.

In the early years of biotechnological research, attention focused on obtaining molecules with new properties, while more recently this has shifted towards optimization of a molecule in relation to its application. Therefore, insight is needed in the requirements for the application, as well as the properties of the biomolecules; this often necessitates a multidisciplinary approach.

Enzymes that catalyze redox reactions are of interest for industrial applications like the (trans)formation of coloring and flavoring agents. The chemistry of such transformations is complex and enzyme mechanism is not always easily understood. Quantum chemical calculations on model complexes can then be helpful for understanding which factors of the redox system contribute to the catalytic process. Until recently the presence of transition metals prohibited the computational treatment of large molecules. And as the interactions of transition metals are difficult to treat properly with simplified classical force fields, semi-empirical methods have also been of limited use. With the recent success of Density Functional Theory (DFT) in treating large systems in an efficient and accurate manner, the calculations have gained importance over the last few years.

One of the projects within the *Computational chemistry of biosystems* program is the *Modeling of metal-based redox reactions*, which involves two PhD-positions (Marieke van den Bosch in Leiden, Marcel Swart in Groningen) under the supervision of the professors G.W. Canters (Universiteit van Leiden), H.J.C. Berendsen and J.G. Snijders (Rijks*universiteit* Groningen). In this project, a combined approach of DFT calculations and Molecular Dynamics (MD) simulations is applied to the copper protein azurin, in an attempt to get a good description of the active site of the protein. The project involves the creation of a copper force field, including the development of two methods to extract force field parameters from the DFT calculations that can be used in the MD simulations; the geometry optimization of the active site in the presence of the protein; the calculation of magnetic, redox and UV/Vis properties that characterize the protein, as well as the application of DFT to two transition metal catalyzed chemical reactions, one of which takes place in a copper-containing enzyme. This thesis is the first one resulting from this project.

Contents

chapter one

INTRODUCTION

1

and historical background

CONTENTS

SUMMARY

As an introduction to this thesis a short description is given of two central topics in this thesis: computational chemistry and copper proteins. When needed further background information will be given in the chapters themselves. In the last section of this Introduction an overview of the contents of the chapters in this thesis is given.

1.1 *Computational chemistry*

Introducing the basic concept

Computational chemistry[1], as the term says, deals with computations that are used to either enhance the understanding of chemical processes or to obtain data that are difficult to acquire experimentally. It can be applied to a variety of fields, but in this thesis it will be applied to biochemical systems. Not only does the field of biochemistry pose a challenge due to the size of the systems involved, it becomes more and more important as our understanding at the experimental level has increased enormously in the past fifty years. A further challenge is posed by the metalloproteins, which are frequently used by Nature as the inclusion of a metal enables vital processes like electron transfer to be executed in an efficient manner. The catalytic activity of the metal atoms is also harnessed by Nature by incorporating them into metalloproteins. From a computational point of view, the metalloproteins pose a real challenge as the interactions of a metal atom with the protein matrix are difficult to generalize in terms of a simplified force field.

Using quantum mechanics to treat the interactions of metal atoms accurately has only recently become possible with the use of Density Functional Theory (DFT)[1]. Although the theory and practical framework were already developed in the 1960's, only in the last fifteen years has this method really gained importance due to the development of new and accurate exchange-correlation potentials. Also the increase in computing power contributed to the range of applicability of the method. Currently, systems with a size of more than a hundred atoms can readily be treated by the method, if appropriate software with efficient parallel computing and linear scaling techniques is being used.

Even though it might already be possible at present to study a small metalloprotein of about a thousand atoms completely by DFT, performing such a task is time-consuming and might show computational bottlenecks that are not present for "small" systems. More importantly, a large portion of valuable time is spent on a very detailed description of regions that can adequately be described by classical interactions, especially if force fields are used that are designed for use on biochemical systems. Therefore, as important as it is to choose the parameters needed for the DFT calculation (choice of the exchange-correlation potential, size of the basis set, accuracy of the numerical integration), it is equally important to choose for which parts of the system a detailed DFT description is needed and which parts can be described classically. This division of the system into a quantum and a classical part forms the core of hybrid QM/MM calculations, which will be described in Section 2.3 and Chapter 9.

The major part of this thesis is concerned however with DFT calculations, either to obtain force field parameters (atomic charges, force constants) for use in classical molecular dynamics simulations (Chapter 3 and 6), application of DFT to a variety of chemical systems (Chapters 4 and 5) or metalloproteins (Chapters 6 to 9). In Chapter 2, a brief description of the methods used in computational chemistry is given, while in the next section (Section 1.2) a short introduction and historical background of metalloproteins is presented.

1.2 *Copper proteins*

Properties, functions and interest

There has been an enormous interest in metalloproteins over the last decades, due to their special properties, abilities and functions in nature[2-63]. A substantial part of these investigations involves proteins where copper is incorporated, since it is one of the most abundant metals found in proteins[64]. Moreover, copper is a multivalent metal, which means that it can occur naturally in its monovalent or *reduced* form (formal charge +1) as well as its divalent or *oxidized* form (formal charge +2), and is therefore extensively used in redox processes. A special class of copper proteins (the so-called blue copper proteins) is characterized by an intense blue color, and a molar absorption coefficient (5500 M^{-1} cm^{-1}) that is several orders of magnitude larger than for molecular model copper complexes. Based on their spectroscopic features (electron paramagnetic resonance (EPR), UV-visible (UV-vis), Resonance Raman (RR))[16], the copper proteins have been classified according to three types:

- type 1 copper proteins contain one copper ion, exhibit an unusual EPR spectrum with a hyperfine splitting appreciably smaller (~0.0060 cm^{-1}) than that found for simple copper complexes (~0.0160 cm^{-1}). They exhibit an intense blue color.
- type 2 copper proteins exhibit EPR spectra similar to those of simple copper complexes.
- type 3 copper proteins contain a dinuclear copper site and usually, as isolated, are EPR silent, which means that the copper atoms are either in the reduced form, or antiferromagnetically coupled. There are no pronounced features in the optical spectrum visible.

Associated with each of these types is a characteristic active site geometry and a characteristic set of metal ligands. For instance in type 1 (or blue) copper proteins like azurin and plastocyanin, the copper is usually bonded to one cysteine and two histidine residues that are located roughly in a plane, with a fourth axial ligand that is in almost all cases a methionine residue or a glutamine. Another important residue is an asparagine (residue 47 in azurin, 38 in plastocyanin), directly preceding the N-terminal histidine ligand (residue 46 or 37), that is conserved in all type 1 copper proteins. One aspect of the positioning of this residue is that it is hydrogen bonded to the residue adjacent to the cysteine ligand (either a threonine or a serine); this fixes the loops with the copper ligands with respect to each other. Another conserved feature is the hydrogen bond of the backbone nitrogen of this asparagine to the sulphur of the cysteine ligand. In many cases a second hydrogen bond to this sulphur is provided by the backbone nitrogen of a closeby phenylalanine residue.

As indicated above, type 1 copper proteins exhibit an intense blue color due to a strong absorption band at ~625 nm with a molar absorption coefficient of 5500 M^{-1} cm^{-1} (corresponding to an oscillator strength of 0.05), which is roughly a factor 100 larger than for normal copper complexes. In the 1960's, when the structure and copper ligands of type 1 proteins were still uncertain, the absorption band at 625 nm was assigned to a d→d transition[65-67], while Williams proposed it to be a ligand to metal charge transfer band[65,68]. Later studies substantiated the ligand to metal transition[69-71], especially when the presence of the thiolate as a ligand had been observed in the crystal structures. However, the charge

transfer assignment has been challenged as it was found that no charge transfer occurs, and the absorption band was assigned to a π-π^* transition.

Azurin

The copper protein under study in this thesis is azurin, a type 1 copper protein that serves as an electron transfer protein. The protein occurs in a variety of bacteria; the most commonly employed azurins are from *Pseudomonas aeruginosa*[3-5,7,59,71-83] (consisting of 128 amino acids) and *Alcaligenes denitrificans*[6,36,42,84-86] (129 amino acids). The active site of azurin is special in that it contains a fifth coordinating group for the copper (Gly45) in an axial position with a Cu-O distance of 2.8-3.0 Å; the others being a cysteine (Cys112) with a Cu-S distance of 2.1-2.3 Å and two histidines (His46, His117) with Cu-N distances of 1.9-2.1 Å in a plane, and an axial methionine with a Cu-S distance of 3.0-3.2 Å (see Figure 1.2.1).

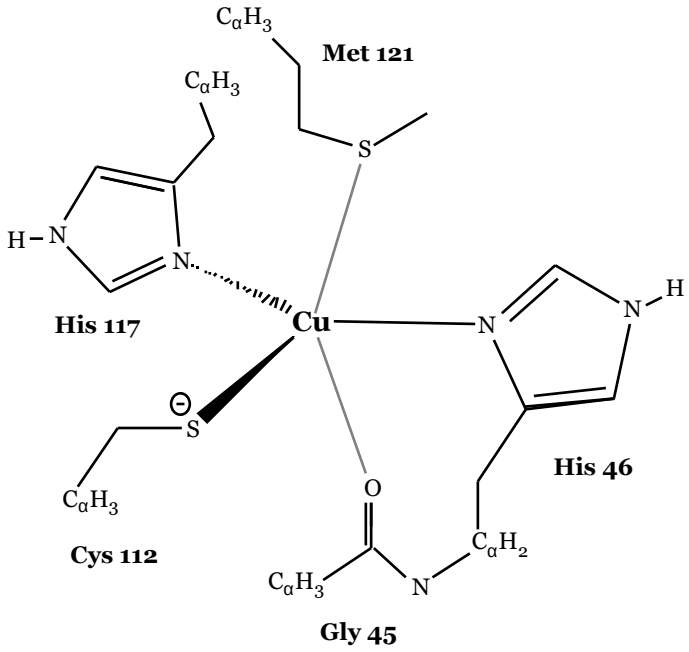


FIGURE 1.2.1. ACTIVE SITE OF AZURIN

The apoprotein structure[79] revealed that there is hardly any change upon the entering of the copper, which results from an extensive hydrogen bond network in the sphere surrounding the copper that provides a rigid protein structure. One of the key residues involved in this network seems to be Asn47[84], which is hydrogen bonded to Ser/Thr113 and connects the loops containing the copper ligands. This rigid structure also limits the change in the active site geometry upon reduction, thereby lowering the energy barrier that has to be overcome in the redox process and enabling fast electron transfer. Another feature commonly found in azurins is a sulphur bridge between residues Cys3 and Cys26 that connect two β-strands. Although this S-S bridge may be important for the stability of the protein, it does not influence the active site structure or characteristics[27,87-89]. The same can be said for the hydrogen bonds of the Asn47 residue towards Ser/Thr113. They can be destroyed as in the case of the Asn47Leu mutant[84], without a significant effect on the spectroscopic features of the protein. However a dramatic decrease in protein stability is observed.[84]

One of the most extensively used techniques to study the function of a particular residue in the protein is site-directed mutagenesis[3,42,75,84,87,90-95], where this particular residue is mutated for another residue. The change in spectroscopic, kinetic or other features, if found at all, then gives a clue to the significance of this residue for the property under study. For instance, the reduction potential is found in the range from 205 to 510 mV depending on the residue at the 121 position[3]. Site-directed mutagenesis can also be used to probe the properties of the mutant in order to compare the spectroscopic features of it with those of another protein for which the structure is not yet solved. This has been successfully applied in the case of the Met121Gln (M121Q) azurin[42] mutant, which was constructed as a possible model for stellacyanin. The spectroscopic features of the M121Q mutant and stellacyanin are very similar and later crystallographic studies[96] confirmed that a glutamine residue was present at the axial position.

Computational studies

Copper proteins have not only received lots of attention from an experimental point of view, they have also been studied in many computational investigations. These comprise molecular dynamics studies[13,21,24,29,44,52,89,97-102] to study the dynamical behavior of the protein, as well as quantum chemical studies[9-14,17,20,34,41,49,57,58,69-71,86,103-107] on model systems of the active sites of these proteins. The former have been used for instance in the calculation of the redox potential difference between wildtype azurin and a mutant, or to investigate the influence of the water molecules on the protein[24]. The latter have been used to study the spectroscopic (EPR, UV-vis) features[34,49,57,58,69,71,86,105-107], to optimize the geometry of the active site model[9-14,17,20,41,103,104] or to check the nature of the bonds between copper and its ligands[107]. In the last study, it was concluded that there is some covalent character in the bond between copper and the axial Met121, while there is hardly any in the bond between copper and Gly45.

In a number of papers, Ryde et al.[9-14,17,20,41,103,104] have reported in detail the geometry optimization of an extensive set of model systems for the active site of type 1 copper proteins, and the corresponding energies in going from the reduced to the oxidized state. They have shown that the active site is not strained, as was predicted by the entatic state theory[22,55,108,109]. This theory states that a group, for instance a metal or part of an amino acid residue, may be forced into an unusual geometric or electronic state by the protein; in the case of the blue copper proteins, both reduced and oxidized copper would be forced into a unfavorable geometry to facilitate electron transfer. However, the computational studies[9-14,17,20,41,103,104] on model systems showed active site geometries that are similar to the crystal structure, while the energy difference between the reduced and oxidized geometries (inner reorganization energy) is small. Still, some problems were encountered in finding the optimal distances for the axial ligands, which were found to be too short (2.3 Å calculated vs. 2.9 Å experimental) or too long. In the case of azurin, it was even impossible to keep the Met121 at an appropriate distance. The inclusion of the protein matrix in the geometry optimization in subsequent QM/MM calculations[17] resulted in better agreement with the experimental structure. Therefore, even though the active site structure can be reasonably represented by model systems, the influence of the protein matrix on it can not be ignored completely, especially for the positioning of the axial ligands. This topic will be discussed in more detail in Section 6.3 and Chapter 9.

1.3 *Thesis overview*

A short description of the contents of this thesis

In this section an overview of this thesis is given, with a short description of the contents of each chapter.

Chapter 2 gives background information on the computational methods used in this thesis, ranging from quantum chemical methods like Density Functional Theory (DFT), via classical mechanics to classical Molecular Dynamics (MD) simulations. Also the concepts of geometry optimizations and hybrid QM/MM, which are frequently used in this thesis, are explained.

In Chapter 3, two methods are presented to obtain force field parameters that are needed in classical mechanics calculations. A new charge analysis is presented that provides atomic charges that reproduce *by construction* the atomic and therefore also the molecular multipole moments, and a new method to obtain from quantum chemical Hessian matrices force constants for the bonding interactions in classical mechanics is discussed.

The application of Density Functional Theory to three different chemical topics is discussed in Chapter 4. First the computation of molecular polarizabilities and the influence of the choice for the basis set and exchange-correlation potential on the accuracy is discussed; then the accuracy of optimized geometries of several exchange-correlation potentials in a number of basis sets is presented, where the test set consists either of a set of small molecules that was used previously by others to check the accuracy of several wavefunction based methods. Finally, the new charge analysis presented in Chapter 3 is validated by checking its use for the concepts of molecular recognition, electron withdrawing-donating groups and electrophilic substitution reactions.

Chapter 5 deals with the application of Density Functional Theory to reaction mechanisms involving metal atoms as a vital and necessary prerequisite. The first part concerns an organic reaction where zinc plays an important role in the aminoalcohol/thiol promoted asymmetric addition of dialkylzincs to aldehydes. In the second part, the mechanism of the reaction taking place in the active site of the copper enzyme quercetinase is studied at all stages of the reaction.

The creation of a copper force field for use in azurin (and derivatives) is described in Chapter 6, including a section concerning the axial bonding in the active site of azurin. The creation of the copper force field involves finding suitable atomic charges for the reduced and oxidized state and at stages inbetween these states. The effect of the electric field of the charges in the surrounding protein on the energy and atomic charges is checked, and the force constants for copper-ligand interactions determined. In the last section, the vibrational frequencies from either DFT, the copper-ligand force constants or the ones obtained in MD simulations are compared.

The calculation of g-tensors and hyperfine coupling tensors of copper complexes and copper proteins by DFT studies is presented in Chapter 7. It is shown that a special procedure is needed in order to get reasonable agreement between computed and experimental g-tensor values. The computed copper hyperfine coupling constants are sometimes in disagreement with the experimental values, but in many cases a good agreement is observed. For the other

atoms, especially in *wildtype* azurin, a good agreement is found between the computed and experimental hyperfine couplings.

Chapter 8 deals with the UV/VIS spectra, or excitation energies, of copper proteins. As Time-Dependent Density Functional Theory (TD-DFT) is formally not applicable to open shell systems and practically not implemented in the ADF program, use is made of the semi-empirical INDO/1 approach. TD-DFT results for the reduced state of three copper proteins are used to calibrate the INDO copper parameters, which are subsequently used for the excitation energies in the oxidized state of these proteins. The influence of the presence of the protein on the computed excitation energies is checked.

Using hybrid QM/MM calculations to optimize the geometry of the active site of metalloproteins in the presence of the protein and a layer of solvent molecules is presented in Chapter 9. A new link model for use in these QM/MM studies is presented that minimizes the influence of the introduction of artificial capping atoms is presented, which is used in subsequent optimizations of the active sites of *wildtype*, *mutated* and *metal-substituted* azurin.

chapter two

Methods

2

CONTENTS

SUMMARY

A background is given on the computational methods used in this thesis, ranging from quantum chemical methods like Density Functional Theory (DFT), via classical mechanics to classical Molecular Dynamics (MD) simulations. Also the concepts of geometry optimizations and hybrid QM/MM, which are frequently used in this thesis, are explained.

2.1 *Quantum chemistry*

Solving Schrödinger's equation

Chemistry deals with molecules and atoms[110]. If one wants to get an accurate description of chemical processes, one should treat the atom by its own constituents: a small nucleus and a large electron cloud. At this point, it is no longer valid to treat the particles in a classical manner, i.e. Newton's classical equations of motion no longer hold. In Section 2.2, we shall return to the classical description and get an indication to what extent the classical picture is valid in providing a reliable description for the system under study.

The fundamental equation which states explicitly why the classical picture no longer holds, is Heisenberg's uncertainty principle:

$$\delta p \ \delta r \geq \tfrac{1}{2}\hbar \tag{1}$$

It simply states that at any point in time, it is not possible in any way to get at the same time both the position and momentum of an electron exactly; the uncertainty of the products is at least as large as the quantity $\hbar$ (Planck's constant, $6.6{\cdot}10^{-34}$ Js, divided by 2π). The fact that this quantity is so small compared to our normal day world is the main reason for to the correctness of Newtonian mechanics for describing macroscopic phenomena. However, the dimensions of an electron are comparable to the size of the atom to which it belongs, and the uncertainty principle can no longer be ignored. Therefore, the electron can no longer be treated as a particle, but rather a wavefunction should be used to indicate the probability that an electron is located at a certain position in space. That is, we enter the field of quantum mechanics[1,111,112] and we have to use the Schrödinger equation:

$$\hat{H}\psi = i\hbar\frac{\partial\psi}{\partial t} \tag{2}$$

This time-dependent equation when solved, produces the wavefunction ψ for a number of electrons that can be used to get a probability distribution for the electrons in space. It also gives the energy associated with that particular state, or electron distribution. The Hamiltonian $\hat{H}$ consists of the following parts:

$$\hat{H} = \hat{T}_e + \hat{T}_N + \hat{V}_{eN} + \hat{V}_{ee} + \hat{V}_{NN} \tag{3}$$

Here, T_e *and* T_N stand for the kinetic energy of the electrons and nuclei respectively, V_{eN} for the nuclear attraction, V_{ee} for the electron-electron repulsion and V_{NN} the nuclear repulsion. Since the difference in mass between the electrons and nuclei is so large (electrons are of the order of magnitude of 10^5 times lighter than nuclei), the Schrödinger equation is normally split up into an electronic and a nuclear part (Born-Oppenheimer approximation). It is then solved only for the electronic part, keeping the positions of the nuclei as external parameters in solving it. Even then, the Schrödinger equation can be solved exactly only for systems containing *one* electron only; for all other systems, approximations have to be used.

The conventional way of solving Schrödinger's equation starts with Hartree-Fock theory[1,111,112]. In this method, electron correlation (interaction between electrons with different spins) is not present explicitly; this correlation has to do with the Coulomb hole. This hole is connected to the interaction of two unlike spins, while the so-called Fermi hole is connected to two like spins. As electrons are Fermi particles, two electrons with the same spin can not be present at the same time at the same position (Fermi hole). Two electrons with differing spin (one alpha, the other beta) *can* be present at the same time at the same place, but are less likely to do so due to electrostatic repulsion (Coulomb hole). As the wavefunction is built up of Slater determinants, the Fermi hole is correctly taken care of. But the Coulomb hole is in the Hartree-Fock theory only taken into account in an average way. The way to improve upon this, is to include determinants where electrons are "promoted" from occupied to unoccupied orbitals, and optimize the weight with which these additional determinants should be added. The two most simple methods to improve upon Hartree-Fock, are Configuration Interaction (CI) and Møller-Plesset perturbation theory. Although these methods usually improve upon Hartree-Fock results, they are not nearly as accurate as more sophisticated methods like Multi-Configurational Self Consistent Field {CASSCF/CASPT2}, Coupled Cluster {CCSD(T)} or Multi-Reference CI. However, the more accurate the method, the higher the computational cost; for instance, while Hartree-Fock scales formally as N^4 (where N is the number of particles), CCSD(T) scales formally as N^7. Therefore, these methods can be applied only to relatively small molecular systems.

A radically different approach is the development of Density Functional Theory (DFT)[1,113]. Already in the early 60's, Hohenberg and Kohn[114] have proven that the ground-state electronic energy is determined completely by the electron density ρ. That is, there exists a one-to-one correspondence between the electron density of a system and the energy. This does not seem that surprising, as indicated by the well-known spectroscopist Bright-Wilson who stood up at a conference in 1965[115], and shared his thoughts about these new ideas with the audience: "*well, the total electron density defines the number of electrons in the system; the cusps in the density define the nuclear coordinates; the derivative of the density at a cusp defines the nuclear charge at that cusp and thus the configuration of the elements in the molecule; therefore, the system is fully defined*". Real life however is not that simple, because although the Hohenberg/Kohn-theorem states that there should exist one density functional that describes the connection between density and energy, it does not provide the functional as such. Even though the fundaments of DFT had already been completed in 1965 with the Kohn-Sham theory[116], the standard use on a large scale emerged at the end of the 80's and beginning of the 90's, when more accurate exchange-correlation potentials had been developed. Eventually, one of the founders of the theory (Walter Kohn) was awarded the Nobel prize for it; at that time (1998) and from then on, DFT has overtaken the use of conventional quantum chemistry due to its efficiency. As it scales formally as N^3, it enables to study larger systems in a reasonable time with an accuracy that could not be reached with for instance Hartree-Fock theory. In recent years, several DFT codes have successfully implemented special techniques[117,118], which make the method scale linearly with the number of particles; this enables one to treat even larger system (up to hundreds or even thousands of atoms) at an accurate level in a reasonable time.

In Density Functional Theory, the point of view changes with respect to the wavefunction based methods. Taking Hartree-Fock theory as an example for the latter, in wavefunction based methods the wavefunction governs everything, and the charge density results from it. In DFT, the picture is reversed. The charge density governs everything and it can be represented by molecular orbitals (the Kohn-Sham orbitals). The Hamiltonian now is built up in another way also; this is coming directly from the Kohn-Sham theory, which states that there exists a system of non-interacting electrons that produces *exactly* the same density as a system of interacting electrons. The kinetic energy T_S of the

non-interacting electrons is known exactly from the molecular orbitals, and forms together with the nuclear attraction energy U_{ne}, the Coulomb energy U_c and the exchange-correlation energy U_{xc} the components of the total DFT energy U_{dft}:

$$U_{dft} = T_s + U_{ne} + U_c + U_{xc} \tag{4}$$

The nuclear attraction energy U_{ne} is the attractive interaction between the nuclei and the electrons and the Coulomb energy U_c the repulsive interaction between electrons; the exchange-correlation energy U_{xc} is the sum of the exchange and correlation energy, as well as a correction term for the kinetic energy for the system of interacting electrons. By equating U_{dft} to the exact energy, this expression can be used to define U_{xc}, which is coming from the "exact" exchange-correlation potential, if it were known. Unfortunately it is not, and approximations have to be made.

Several exchange-correlation potentials are currently available. The simplest of them is the Local Density Approximation (LDA), that results directly from the description of the uniform electron gas. The most commonly found implementation for this potential is represented by Vosko, Wilk and Nusair (VWN)[119]. However, as this uses only the local density and as such underestimates the interactions due to other atoms, the generalized gradient approximations have been introduced in the mid 80's. They take also the derivative of the density into account, and generally give a better description for a wide range of phenomena than LDA. The most popular *pure* DFT potential is probably the combination of the Becke exchange[120] and the Perdew correlation[121] potentials, which has successfully been applied to predicting geometries and other properties, and was constructed on physical grounds.

However, it can not be left unnoticed that the most popular DFT-potential at this point in time is the B3LYP potential[122]. It combines *pure* DFT potentials like BLYP with a portion of exact Hartree-Fock (HF) exchange, where the amount of mixing is based on empirical grounds, i.e. fitted to improve some properties for a certain set of molecules. Although the physical motivation for this choice may be lacking, it is shown that this combination improves the performance of the BLYP potential (see also Section 4.2), but gives results comparable to the *pure* Becke-Perdew potential. Recently, both the B3LYP and the Becke-Perdew potentials have been improved (by the Becke97[123] and the PKZB potential[124,125] respectively), where again the former was based on fitting to some properties of some molecules. The PKZB potential instead is one of the Meta-GGA's where not only the first but also the second density derivative is used, and has been constructed again on physical grounds. Therefore, if one is interested in the properties the B3LYP or Becke97 was optimized for, and one wants to apply it to a molecule similar to the ones in the set used for fitting the parameters, it is probably safe to use them. However, it seems more reasonable to study the molecule with *pure* DFT potentials as these give not only comparable accuracies for the properties the B3LYP/Becke97 were optimized for, but also provide a better description for properties or molecules that were not used in the B3LYP/Becke97 set for obtaining the parameters.

To get an idea of the accuracy that can be obtained with DFT with currently available exchange-correlation potentials, one is referred to Chapters 4 and 5, where DFT has been applied to study several chemically interesting properties or systems, like for instance molecular polarizabilities of organic molecules, geometries of molecules or even the complete reaction path of an aminothiol/aminoalcohol catalyzed addition of dialkylzinc to benzaldehyde.

2.2 *Classical mechanics*

Treating interactions at the atomic level

In Section 2.1, the electrons and nuclei were the particles around which everything evolves. In this section, the interactions are treated at the atomic level. That is, instead of solving Schrödinger's equation for obtaining the energy, the interactions are given by classical energy expressions, or force fields. Many force fields are available, each with their own set of force field parameters that can be applied to either a certain class of molecules (for instance proteins or organic molecules) or to any molecule, that can be built from any atoms. Example of the former are the AMBER95[126] and GROMOS96[127] force field, that were designed for applications on biochemical systems like proteins; the latter are exemplified by for instance the Universal Force Field (UFF)[128] or the MM2/MM3[129] force fields. The interactions within any force field can normally be divided into two parts: bonding and non-bonding.

Bonding potentials

These potentials deal with bonding interactions between close-by atoms, that are normally connected by a chemical bond; as such they replace for a large part the interactions which should normally be treated at the quantum chemical level. The bonding interactions are a function of bonds, angles and dihedral angles between different atoms. The simplest way to describe the interactions for these *internal* coordinates is by the harmonic approximation:

$$U(X) = \tfrac{1}{2} K_X \left(X - X_{eq} \right)^2 \qquad (1)$$

In this equation, the energy changes quadratically as the *internal* coordinate changes from its equilibrium value. The $\boldsymbol{K}_X$ parameter is called the force constant and determines how much the energy changes if the *internal* coordinate changes; a small value defines a flexible coordinate, while a large value indicates a rather stiff coordinate. In Section 3.2, a new method (IntraFF) is introduced that can be used to obtain values for force constants that are not present in the force field one is using. There are of course many other ways to treat the bonding interactions; for instance, in the GROMOS96[127] force field a quartic potential is used instead of the harmonic potential. It was introduced to avoid the square root operation in the calculation of the energy and the forces working on the atoms. In practice however, the difference between the quartic and the harmonic potential is rather small; the differences are substantial only at larger deviations from the equilibrium value, which are in practice never reached in the calculations.

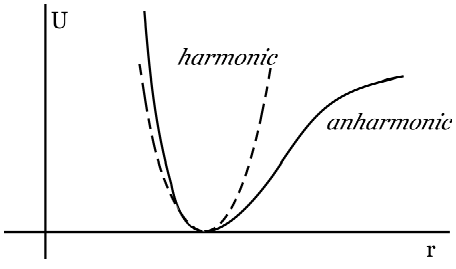


FIGURE 2.2.1 ANHARMONIC POTENTIALS

Other potentials that differ to a large extent from the harmonic potential are the Morse[130] or Frost[131,132] potentials; both are anharmonic potentials for bonds, and give an improved description of the energy profile as the bond is stretched. That is, if the atoms making up the bond are pushed close to each other, the energy increases enormously, while if they are stretched really far apart, eventually the system is treated as two non-interacting free atoms, as would be expected. The well-known Morse potential has three parameters: the equilibrium distance, the dissociation energy and the anharmonic constant; the less well known Frost potential only has two, and is based on theoretical grounds. Both methods are described in more detail in Section 3.2, where it is shown also how to obtain the parameters for the Frost potential.

Dihedral potential

For the dihedral angles, one normally uses a periodic function with minima at either 120, 180 or 240 degrees apart (depending on the number of atoms connected to the central atoms on either side). The most convenient way to treat this situation is by assuming a (combination of) cosine potential(s) for the dihedral angle connecting the four atoms involved:

$$U_{ijkl}^{dihedral} = K_{\phi}\left[1+\cos\left(n\phi - \phi_{shift}\right)\right] \tag{2}$$

This assumes that there is no preferential minimum involved; i.e., it is based on the case of ethane, which has a favorable (*staggered*) and an unfavorable (*eclipsed*) conformation, while there is no preference for two hydrogen atoms to have a dihedral angle of either 60, 180 or 300°. However, this is in general not valid, but a properly chosen combination of dihedral angle functions will give a proper representation of the potential energy surface again.

Non-bonding interactions in non-polarizable force fields

The interactions between molecules that are not involved directly in bonding interactions, are called the non-bonding interactions, and comprise normally electrostatic and van der Waals interactions. The former is represented by the interactions of point charges, which should be properly chosen to give a good representation of the charge distribution within a certain molecule. The latter interactions are in non-polarizable force fields normally represented by a Lennard-Jones potential, which consists of a short range repulsion term and a longer range attraction term (due to dispersion terms):

$$U^{Lennard-Jones} = \frac{C_{12}}{r^{12}} - \frac{C_6}{r^6} \tag{3}$$

Each of the force fields usually has its own set of Lennard-Jones (or vdW) parameters, where the particular choices made for them depend to a large extent on the parameters chosen for other parts of the force field. For instance, as polarization effects are absent in these force fields, for several force fields it was chosen to mimic these effects by increasing the charges on the atoms. Other force fields have been developed without such an increase, and the effect of polarization interactions has been introduced in an effective way by changing the vdW parameters.

Electrostatic interactions are a special kind of interaction, as they have has a long range influence; two unit point charges at a distance of 100 Bohr (~53 Angstroms) still have an electrostatic interaction of approximately 6 kcal/mol! This is of the same order of magnitude as van der Waals interactions between two atoms that are very close (~3 Å). Therefore, treating these interactions properly is of the utmost importance; fortunately, this large interaction between distant atoms is only present if the

atoms are isolated. In condensed matter, the electrostatic interactions are screened by the atoms inbetween, which reduce the interactions by a factor of 2 to 80, depending on the kind of matter.

Non-bonding interactions in a polarizable force field

Polarization effects arise when two molecules approach each other and start to feel the presence of the other molecule; depending on the particular orientation of the two molecules and their respective electrostatic molecular multipole moments, the charge distribution within each of the two molecules will start to differ from their monomeric distributions. When the system of the two molecules are treated with quantum chemical methods, these polarization effects are treated properly. In a non-polarizable force field, these effects can be treated only in an effective way, as described earlier.

The polarization effects can however be treated properly also at a classical level by using the (*atomic* or *molecular*) polarizability. This property is related to the change in dipole moment of the molecule in an electrostatic field; this field can then be either an external macroscopic field or the field of a neigboring molecule with permanent electrostatic multipole moments. It can be even the field at molecule A of the dipoles induced in molecule B by the field of the permanent multipole moments of molecule A!

For a set of atomic polarizabilities in an electrostatic field, the *induced* dipole moment in each of the polarizable points *p* or *q* is given by the polarizability times the total field; this total field consists of the electrostatic field E_O (from all the permanent dipole moments and/or an external electrostatic field) plus the dipole field T_{pq} of all other *induced* dipoles:

$$\mu_{p,induced} = \alpha_p \left(E_{static} + \sum_{q \neq p} T_{pq} \mu_{q,induced} \right) = \alpha_p E_{total} \qquad (4)$$

This can be rewritten in a matrix formulation, leading to a matrix that should be inverted to get the relay matrix (see also Section 4.1). This *3Nx3N* relay matrix then gives the response of the molecule to a given electrostatic field, and as such represents the molecular polarizability (in a *3Nx3N* representation). It can be reduced to a *3x3* form to retain the normal molecular polarizability tensor (see Section 4.1).

In the Direct Reaction Field (DRF) approach[133-142], extensive use has been made of this polarizable nature of molecules, in combination with a screening function that accounts for overlapping charge densities. In principle, the approach has been primarily used for QM/MM purposes (see Section 2.3), but also a classical force field has been developed. In this polarizable force field, the non-bonding interactions are split up into six terms: *electrostatic*, *dispersion*, *repulsion*, *induction*, *field energy* and *long-range electrostatics* (through either a dielectric continuum description or an image charge method; in both cases, the Poisson-Boltzmann equation is solved). The *electrostatic* interactions are similar to the ones used in non-polarizable force fields, except that the interactions are screened for overlapping charge densities. The *repulsion* and *dispersion* interactions can be thought of as similar to the Lennard-Jones potential, except that again the dispersion interactions are in the DRF approach screened for overlapping charge densities, and it is normally obtained as the anisotropic interaction between two polarizability tensors. The Lennard-Jones treatment of isotropic interactions is generally speaking only valid for isolated atoms, not for atoms within a molecule. Moreover, unlike most non-polarizable force fields, the parameters needed for describing the interactions are coming directly from properties of the isolated molecule itself (see also Section 4.1). The *induction* interactions are obtained directly from the interaction of the polarizabilities with the electrostatic field:

$$U^{induction} = -\frac{1}{2} E_{static} \cdot \mu_{induced} = -\frac{1}{2} E_{static} \cdot E_{total} \quad (5)$$

The induced dipole moments can be obtained either from the relay matrix times the electrostatic fields, or from an iterative scheme. The former needs a matrix inversion of a matrix of size $3Nx3N$, which takes a long time, especially for larger systems; therefore, the iterative scheme is much preferred, as it is much faster and due to the form of the relay matrix, extremely stable. The *field energy* treats the direct interaction of a permanent molecular moment with an external field, and is only present if an external field is present. Finally, the *long range electrostatic interactions* can be treated in two different ways, which both invoke a dielectric continuum environment of the system. The first way to deal with this is by employing an enveloping surface of Boundary Elements[133,135-138,143-145], representing a solution to the Poisson-Boltzmann equation. However, this method can be applied only to moderately sized systems (like a solvated organic molecule, not for an entire protein), as the size of the relay matrix is of the order of $3xN+N_{bem}$, where N stands for the number of polarizable points and N_{bem} for the number of Boundary Elements. A more convenient solution to the Poisson-Boltzmann equation can then be obtained by employing Friedman's image charge method[146], or 3D finite difference methods[147].

In the Direct Reaction Field approach[133-142], the correct physics is employed at all levels. Going from the quantum chemical system to the classical polarizable system to the dielectric continuum, everything is consistently coupled. In principle one can treat two atoms either by a full quantum chemical description, a mixed QM/MM description (see Section 2.3) putting either one of the two in the classical system, or both in the classical system, and end up with the same total picture for the interaction, energy and charge distribution. However, apart from some isolated instances, the approach has not been applied systematically to describe the interactions within proteins, which is one of the major goals for the near future.

2.3 *Hybrid QM/MM methods*

Combining quantumchemical and classical systems

In Section 2.1, the energy was obtained for a pure quantum chemical system, while in Section 2.2 it was obtained for a purely classical system. However, in many cases it is more advantageous to have a system that is described partially quantum chemically and partially classically. In order to do this, one has to divide the total system into two subsystems: a QM system that is described by quantum chemistry and a MM system that is described by a classical description. The interactions between these two systems are then usually treated at the same level as the interactions in the MM system itself.

One of the distinguishing features between the different ways of combining QM and MM systems is concerned with polarization effects. One would like to have a system that can be polarized in all regions, like in quantum chemical calculations. However, for regions described by a non-polarizable force field this is not possible. They can polarize the QM regions, but not be polarized themselves. As the force field parameters are obtained for use in a non-polarizable (MM) system, it is more appropriate to treat the interactions between the QM and MM system at the same level. Therefore, it is better to treat the electrostatic interaction at the classical level and not let the QM system be polarized by the MM system. In the application of a non-polarizable force field in QM/MM calculations on copper proteins (see Chapter 9), the QM and MM systems are therefore coupled through classical (non-polarizable) electrostatic interactions.

For a polarizable force field like the Direct Reaction Field[133-142] approach, the opposite is true, as the MM system is polarized already by itself, and as such should also polarize the QM system. In the past, this approach has been primarily used for solvent effects on organic molecules, reactions, excited state lifetimes and charge distributions, or excitation energies. In those applications, wavefunction based methods have been used for the QM description; progress has been made to couple the Direct Reaction Field approach also with Density Functional Theory[148].

A very important issue related to QM/MM calculations is the treatment of the QM/MM boundary region[149]. For solvent effects on organic molecules, the division in a QM and a MM system is straightforward and doesn't cause any problems. However, for a protein this no longer holds; in order to make a division in a QM and a MM system, one has to cut through covalent bonds (see Figure 2.3.1).

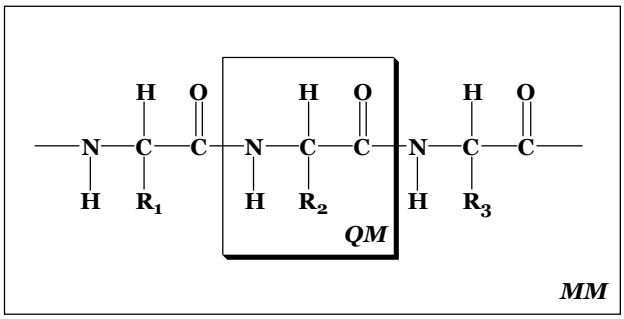


FIGURE 2.3.1. CUTTING THROUGH BONDS IN PROTEINS IN QM/MM SYSTEMS

For instance, if amino acid residue 2 (consisting of $NHCHR_2CO$) is supposed to be put in the QM system and residues 1 and 3 in the MM system, residue 2 is left with dangling bonds. For the MM

system this poses no problem, as the interaction with the QM system is treated on a MM level in which the QM system can be treated as if it were a MM system.

Several options are available to circumvent the problem with the dangling bonds of the QM system (see ref 149). One option, that is preferred if one uses plane-wave basissets, is to use pseudo-potentials[150]. However, for regular QM calculations with atomic basisfunctions this is less straightforward to implement. Another solution is to use localized orbitals, but the most commonly chosen way to solve the problem is to use link atoms. In this method, *capping* (link) atoms are added to the QM system in order to fulfil the valency of the system, which are normally chosen to be hydrogens. The QM calculation is then performed on the *capped* QM system, while the *capping* atoms are not involved in the interaction between the QM and MM systems. In the first paper describing a study with link atoms, no constraint was put on the position of the *capping* atoms, thereby leading to an increased number of degrees of freedom for the total system. In later implementations (like ONIOM[151-153], QMPot[154] or IMOMM/ADF[155-157]), these unwanted additional degrees of freedom were removed by putting constraints on the positions of the atoms involved in such link bonds. Although the number of degrees of freedom is corrected, sometimes this is achieved by removing them for the wrong atom. For instance in the IMOMM/ADF[155-157] implementation, the real classical atom is "following" the *capping* atom; therefore, the degrees of freedom of the artificial *capping* atom are kept, while those of the real classical atom are removed. In Section 9.1, a new link model (AddRemove[158]) is presented that doesn't have this unwanted behavior; furthermore, the introduction of the artificial *capping* atom is corrected for afterwards, and in principle its introduction should have no influence.

In Chapter 9, QM/MM calculations will be applied to obtain active site geometries of copper proteins in the presence of the complete protein and a layer of solvent molecules surrounding the protein. As the computer resources are not unlimited, treating the complete protein by Density Functional Theory is *at present* out of reach. Moreover, by doing so one would waste valuable computer resources as one uses a high-quality description for a large region that can be treated very well by classical interactions. Therefore, only the site of interest (the active site) has been placed in the QM system, while the rest of the protein and the solvent have been put in the MM system. It is shown that including the protein is necessary for obtaining correct active site geometries.

2.4 Optimizations and simulations

Obtaining equilibrium geometries and sampling the degrees of freedom of the system

In the previous sections, only the energies were discussed but not it was not shown how the geometries needed for the calculations can be obtained or improved. One can imagine that the geometry may have a large influence on the energy and molecular properties like the multipole moments. Also the dynamical behavior of molecules at room temperature was not discussed yet. In this section, the basics for methods involved in these two important matters will be discussed, while the applications of the methods are given in later chapters.

Geometry optimizations

The energy profile for a molecule can be visualized as in Figure 2.4.1, where the energy is given as function of some intrinsic coordinate of the molecule. This intrinsic coordinate consists of a combination of all relevant coordinates of the molecule, but can be visualized most easily for a diatomic molecule. In that case, the intrinsic coordinate consists of the distance between the two atoms.

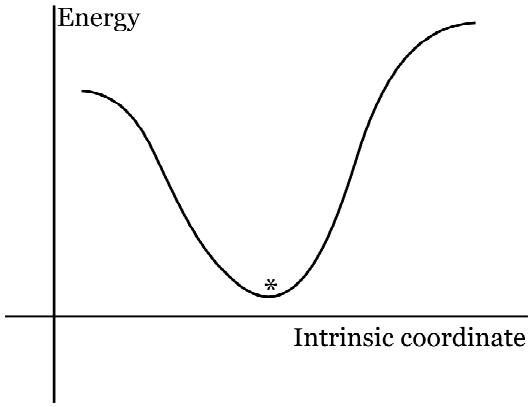


FIGURE 2.4.1. ENERGY PROFILE FOR A MOLECULE

The equilibrium geometry, indicated in Figure 2.4.1 by the star, is found where the energy is lowest. In a geometry optimization, one aims to find this equilibrium geometry. One could use energy values only, but as this requires lots of energy evaluations it is useful only for methods in which the evaluation of the energy does not take a long time (as in the case of semi-empirical or molecular mechanics methods).

A more efficient approach is to use also the gradient g, which is a vector containing the first derivatives of the energy U with respect to the Cartesian coordinates of the atoms $\boldsymbol{r}_i$:

$$\mathbf{g}_i = \frac{\partial U}{\partial \mathbf{r}_i} \tag{1}$$

In Figure 2.4.2, the gradient is sketched that belongs to the energy profile of Figure 2.4.1 as function of the intrinsic coordinate. In the equilibrium geometry, it is zero.

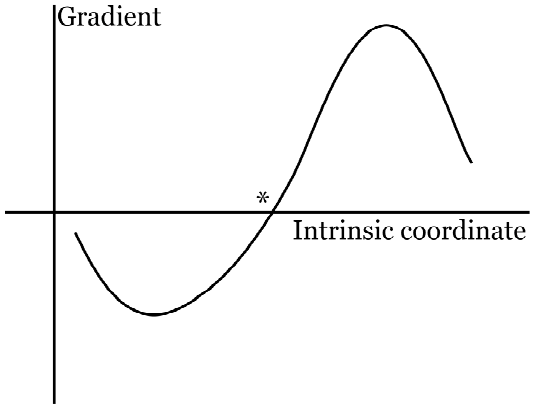


FIGURE 2.4.2. GRADIENT PROFILE

On both sides of the equilibrium, a force is working on the molecule to push it back towards equilibrium. This can be used in geometry optimizations to obtain the equilibrium geometry; one has to find the geometry where the gradient is zero.

The most simple algorithm takes the negative gradient as step in the geometry optimization (*steepest descent*)[1]. This requires lots of optimization steps as the convergence is usually slow. It can be improved upon by using the *conjugate gradient* method. This method constructs its steps by taking a mixture of the current negative gradient and the previous search direction, thereby making it *conjugate* to the previous step. This will improve the convergence to some extent, but the convergence remains slow.

A further improvement of optimization techniques can be obtained by including the Hessian H, which is a matrix containing the second derivatives of the energy U with respect to the atomic coordinates r_i:

$$H_{ij} = \frac{\partial^2 U}{\partial r_i \partial r_j} \tag{2}$$

One could compute this matrix after every geometry optimization step and use it to construct a new step in the optimization, but as the calculation of the Hessian matrix may be very time-consuming, the use of approximate Hessians may be worthwhile. These approximate Hessians are not re-calculated after every geometry optimization cycle, but simply *updated* based on the gradient change. One of the most popular and powerful update schemes is the Broyden-Fletcher-Goldfarb-Shanno (BFGS)[1] scheme. In this scheme, the Hessian is updated from H_0 to H_+ by using the step vector s and the gradient difference Δg:

$$H_+ = H_0 + \frac{\Delta g \quad \Delta g}{\Delta g \cdot s} - \frac{(H_0 \cdot s) \quad (H_0 \cdot s)}{s \cdot H_0 \cdot s} \tag{3}$$

In this equation, the tensor product has been used:

$$(a \quad b)_{ij} = a_i b_j \tag{4}$$

The BFGS scheme has, under certain weak conditions on the step vector *s*, the property of non-negative definiteness. This property is useful as it ensures that the quadratic model has a minimum, or put otherwise, that the eigenvalues of the Hessian are all larger than or equal to zero.

The BFGS is less appropriate for use in a *transition state* optimization. This *transition state* (‡) is defined as the highest point on the energy curve along a reaction coordinate connecting reactants *R* and products *P* (see Figure 2.4.3).

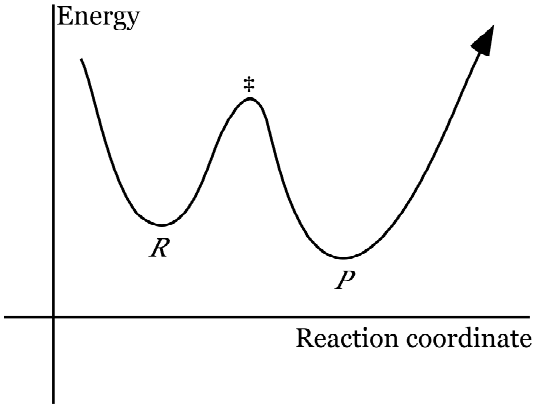


FIGURE 2.4.3. ENERGY PROFILE ALONG A REACTION COORDINATE

At the *transition state*, the energy goes downhill if a step is taken along the reaction coordinate and uphill for all other directions. It is therefore characterized by a Hessian that has one *and only one* negative eigenvalue. A Hessian update scheme that uses non-negative definiteness is therefore not appropriate for finding such a transition structure. The *Powell-symmetric-Broyden* (PSB)[1] update scheme has, like the BFGS scheme, the property that if H_0 is symmetric then also H_+ is. However, the PSB updated Hessian is not positive definite and is therefore better suited for *transition state* optimizations. The updated Hessian is obtained as:

$$H_+ = H_0 + \frac{(s \cdot s)[T \quad s] + (s \cdot s)[s \quad T] - (T \cdot s)[s \quad s]}{(s \cdot s)^2} \tag{5}$$

with the *T*-vector:

$$T = -g - H \cdot s \tag{6}$$

For the geometry optimizations described in this thesis, the BFGS update scheme was used for *equilibrium structure* optimizations of molecules and the PSB scheme for *transition structures*. In Section 4.2, accuracies of optimized geometries obtained by Density Functional Theory are discussed.

Molecular simulations

The statistical or dynamical behavior of molecules at room temperature can have some marked effects on properties one is interested in; thermodynamic properties like equilibria, redox potentials and protein stabilities are determined by probability distributions at a certain temperature, while for dynamic properties like reaction rates also time correlation functions play a role. To enable a comparison between calculations and experiments, one therefore has to do the calculations at a certain temperature and follow the (dynamics of the) system for a certain time. The time scale on which the

system should be followed depends on the property one is interested in. This thesis deals with *atomic* and *molecular* properties; therefore, the dynamical behavior should be handled by using Molecular Dynamics[159-161].

In Molecular Dynamics simulations, the system is followed for a certain time, in which the coordinates of the atoms/molecules are updated at regular intervals (with a timestep that is normally of the order of 1 fs). At every timestep, the energy and the forces working on the atoms are calculated, and the atomic velocities and coordinates updated. Several properties like the energy, the pressure or total dipole moment of the system may be monitored during the simulation, which results in statistical averages for these properties during the simulations. An appropriate total simulation time depends on the system and on the properties one is interested in. For solvent effects on organic reactions, a total simulation time of 50-100 ps seems appropriate, while for the stability check of a protein a few ns may be more appropriate. The simulations generate a certain ensemble, which is determined by the system properties that are constrained during the simulation. In the *canonical* (NVT) ensemble, the total number of atoms, the volume and the temperature are kept constant, while in the *isothermal-isobaric* (NpT) ensemble the pressure is constrained instead of the volume. Several methods have been proposed to keep the temperature (and/or pressure) constant. It is achieved for instance by extending the Hamiltonian (Nosé-Hoover[162-165] dynamics) or by velocity/pressure scaling (Berendsen[166] dynamics). Also, there are several schemes available for updating the atomic/molecular coordinates, the simplest of which is the Verlet scheme. More detailed information about the art of Molecular Dynamics simulations can be obtained elsewhere[159-161].

In Section 6.4, Molecular Dynamics simulations are used to check the force field that is obtained for the active site of Azurin in Sections 6.1 and 6.2. Vibrational frequencies from either Density Functional Theory, the copper force field or MD simulations are compared with each other and with experimental data.

chapter three

FORCE FIELD PARAMETERS

from quantum chemical calculations

3

CONTENTS

SUMMARY

Two methods are presented to obtain force field parameters that are needed in classical mechanics calculations. A new charge analysis is presented that provides atomic charges that reproduce by construction the atomic and therefore also the molecular multipole moments, and a new method to obtain from quantum chemical Hessian matrices force constants for the bonding interactions in classical mechanics is discussed.

3.1 *Multipole derived charge analysis*

A charge analysis derived from an atomic multipole expansion

The electrostatic potential from the charge density $\rho(r_i)$ of a molecule is obtained in a point r_s as[a]:

$$V_C(r_s) = \int \frac{\rho(r_i)}{|r_s - r_i|} dr_i \qquad (1)$$

However, this equation is not well suited for standard use in molecular simulations. One could simplify the integral by using a one-center multipole expansion, but this can not be applied for larger molecules, since one needs to go to high order multipoles. Associated with it is a large convergence radius, outside which the expansion is valid. A better method (for instance the Distributed Multipole Analysis[167,168]) is to use multipoles located at several centers (usually the atoms) with two advantages: the order of multipoles needed is smaller and the convergence radius is per center (and smaller). However, even these methods are cumbersome to use in simulations: all multipole moments should be transformed from the local (i.e. the one in which they were obtained) to the global coordinate system (i.e. the one in the simulation), which involves a number of subsequent [3x3]-transformations for each moment (1 for dipoles, 2 for quadrupoles, etc.). A further disadvantage is that in Molecular Dynamics simulations for the calculation of the forces the derivative of the potential is needed, making it even more cumbersome.

A further simplification is therefore worthwhile: taking only (atomic) charges into consideration. This reduces not only the number of interactions per pair of centers (from 10-35 to 1, depending on the order of multipoles located at the centers), but removes also the need for transforming the moments. The potential is then obtained from the charges q_i at positions r_i as:

$$V_C(r_s) = \sum_i \frac{q_i}{|r_s - r_i|} \qquad (2)$$

Therefore, assigning charges to atoms is an important problem in computational chemistry[1], especially in view of constructing *accurate* force fields for simulations using Molecular Dynamics or Monte Carlo techniques[160,161]. Our interest is in the field of biochemistry, in particular copper proteins[8], where the standard force field parameters, if present, are not accurate enough to model the chemical processes going on in the active site of the proteins. This is especially true for the study of properties like the redox potential, where one wants to see the behavior of the protein when an electron is removed from the system.

Many methods to obtain atomic charges from a quantum chemical calculation are available (detailed surveys can be found elsewhere[1,169,170]), where there is a general preference in the molecular modeling world for *Potential Derived* (PD) charges[1,171]. However, associated with this class of charge analyses are several major drawbacks, all of which are related more or less to the need of a grid of points where the quantum chemical potential is fitted by the electrostatic potential from the charges. It results in:

– a strong dependence on the choices made for determining the grid:

[a] In atomic units (see appendix A.1 at the end of the thesis)

  – an arbitrary change in parameters results in different atomic charges
  – some methods are even orientation-dependent[172,173]
– a strong dependence on the method how to fit the charges:
  – underdetermined set of equations, multiple "solutions" possible
– numerical (in)stability
– great uncertainties in assigning charges to atoms *buried* within the molecule, since the potential outside the molecule is mainly determined by the atoms near the surface[1]
– an enormous increase in CPU-time, since lots of 1/r terms resulting from charge-gridpoint distances should be evaluated many times

Several authors have extended these PD-methods by introducing constraints on the atomic charges; usually the *molecular* multipole moments up to some order should be conserved. This seems to make the methods less grid-dependent, but they are still limited to a relatively small number of atoms (approx. 20; depending on the order of the multipoles[171]) and still suffer from the other drawbacks.

We present here a new charge analysis based on ideas used for the *Dipole Preserving Charge analysis*[174], but formulated in another way and using more accurate *atomic* multipoles. There are three stages involved in this method: we start by writing the molecular charge density as a sum of atomic densities. Subsequently, from these atomic densities a set of atomic multipoles can be defined, which can be used to get the electrostatic potential outside the charge distribution. Finally, these atomic multipoles are reconstructed by using a scheme that distributes charges over all atoms to reproduce these multipoles *exactly*. Therefore, this method does not suffer from the drawbacks of the PD-methods.

An important feature to notice here further, is the advantage of using an *atomic* over *molecular* multipole expansion. Close to any atom the electrostatic potential is mainly determined by the charge distribution around that atom; or within the *atomic* multipole expansion, by the *atomic* multipoles near to that point. This is one of the main advantages over using molecular multipoles only, when one needs to go to high orders to get a good representation of the potential in that point (if at all). Moreover, since the *molecular* multipole moments (up to order X) are reproduced necessarily by the *atomic* multipoles (up to order X), our charges have the nice feature that they do not only represent the *atomic* but also the *molecular* multipoles.

This section consists of three parts: the first deals with how to get from a molecular density to a set of atomic multipoles, in the second we explain the distribution scheme, while in the third, we give some results of the method.

Atomic multipoles from a multipole expansion

A molecular charge density ρ is usually obtained in a basis set expansion (with atom indices A,B, basis functions ϕ_j, basis function indices i,j and P_{ij}^{AB} element ij of the density matrix):

$$\rho = \sum_{AB} \rho_{AB} = \sum_{ABij} P_{ij}^{AB} \phi_i^A \phi_j^B \tag{3}$$

which defines the density as a sum of atom pair densities ρ_{AB}. Next, we fit these atom pair densities by using atomic functions f_i (without specific details about the functions to keep the discussion general; our choice and motivation will be specified in the **ADF implementation** part):

$$\tilde{\rho}_{AB} = \sum_i d_i^A f_i^A + \sum_j d_j^B f_j^B \tag{4}$$

where the coefficients can be obtained from minimization of the density differences:

$$\int \left(\rho_{AB} - \rho_{AB}\right)^2 d\tau \tag{5}$$

The total density can now be written as a sum of atomic densities:

$$\tilde{\rho} = \sum_{AB} \tilde{\rho}_{AB} = \sum_{AB} \left(\sum_i d_i^A f_i^A + \sum_j d_j^B f_j^B \right) = \sum_{Ai} c_i^A f_i^A = \sum_A \tilde{\rho}_A \tag{6}$$

Furthermore, using these functions f_i, the electrostatic potential in a point **s** is given by

$$V_C(\mathbf{r}_s) = \sum_A V_C^A(\mathbf{r}_s) = \sum_A \sum_{i \in A} c_i^A \int \frac{f_i^A(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_s|} d\mathbf{r} \tag{7}$$

Next, an atomic multipole expansion of the r^{-1} term can be used:

$$V_C^A(\mathbf{r}_s) = \sum_l \sum_{m=-l}^{m=l} \frac{4\pi}{2l+1} \frac{M_{lm}^A Z_{lm}(\hat{\mathbf{R}}_{sA})}{\mathbf{R}_{sA}^{l+1}} \tag{8}$$

with the real spherical harmonics Z_{lm} [175] and the multipole "moments" M_{lm}^A where R_{sA} is the distance vector from nucleus A to a point r_s. These "moments" can be obtained from the coefficients as:

$$M_{lm}^A = \sum_{i \in A} c_i^A \int f_i^A(\mathbf{r}_2) \mathbf{r}_2^{\,l} Z_{lm}(\mathbf{r}_2) d\tau_2 \tag{9}$$

These multipole "moments" are not equal to the multipoles from the Buckingham convention (see **Computational details**), but they can be easily transformed into those by inserting the Cartesian expressions for the Z_{lm}'s.

Atomic multipole derived charge analysis

We start from the *atomic* multipoles as they are obtained from the multipole expansion. Then for each atom, we reconstruct the set of multipoles (up to some order X) located on that atom by *redistributed* charges $q_{s,A}$ on *all* atoms that preserve (up to order X) that particular set of multipoles. The multipole moments represented by the charges $q_{s,A}$ are obtained with the position vectors relative to atom A, $r_{is,A}$, as:

$$\begin{aligned}
Q_A^{repr} &= \sum_s q_{s,A} \\
\mu_{i,A}^{repr} &= \sum_s q_{s,A} r_{is,A} \\
\Theta_{jk,A}^{repr} &= \sum_s q_{s,A} \left(\tfrac{3}{2} r_{js,A} r_{ks,A} - \tfrac{1}{2} \delta_{jk} r_{s,A}^2 \right) \\
\Omega_{lmn,A}^{repr} &= \sum_s q_{s,A} \left(\tfrac{5}{2} r_{ls,A} r_{ms,A} r_{ns,A} - \tfrac{1}{2} \delta_{lm} r_{ns,A} r_{s,A}^2 - \tfrac{1}{2} \delta_{nl} r_{ms,A} r_{s,A}^2 - \tfrac{1}{2} \delta_{mn} r_{ls,A} r_{s,A}^2 \right)
\end{aligned} \tag{10}$$

When the number of atoms is larger than the total number of multipole moments (per atom) to be reconstructed, there is of course more than one way to distribute the charges. We therefore use a weight function that falls off rapidly to keep the *atomic* multipoles as local as possible, i.e. as close as possible to the atom where the multipoles are located:

$$w_s = \exp\left(-\frac{|r_s - r_A|}{d_A}\right) \tag{11}$$

where d_A is the distance from atom A to its nearest neighbor, α an exponential prefactor and w_s the weight for atom *s*, when distributing the multipole moments of atom *A*.

Now, we want the *redistributed* charges to be as small as possible, and at the same time constrain the represented multipoles (Q^{repr}, μ^{repr}, Θ^{repr}, Ω^{repr}) to be equal to the *atomic* multipoles from the multipole expansion (Q^{mpe}, μ^{mpe}, Θ^{mpe}, Ω^{mpe}). We achieve this by minimizing the following function where the constraints are met by using Lagrangian multipliers λ_A, $\lambda_{i,A}$, $\lambda_{jk,A}$, $\lambda_{lmn,A}$:

$$\begin{aligned} g_A = &\sum_s \frac{q_{s,A}^2}{2w_{s,A}} + \lambda_A\left(Q_A^{mpe} - Q_A^{repr}\right) + \sum_i \lambda_{i,A}\left(\mu_{i,A}^{mpe} - \mu_{i,A}^{repr}\right) \\ &+ \sum_{jk} \lambda_{jk,A}\left(\Theta_{jk,A}^{mpe} - \Theta_{jk,A}^{repr}\right) + \sum_{lmn} \lambda_{lmn,A}\left(\Omega_{lmn,A}^{mpe} - \Omega_{lmn,A}^{repr}\right) \end{aligned} \tag{12}$$

With this choice of function, we ensure that the distribution mainly takes place close to the atom A where the multipoles are located. After taking the derivative of this function with respect to the redistributed charges $q_{s,A}$ to find the optimal choice for the charges, the following equation is obtained:

$$q_{s,A} = w_{s,A}\left[\lambda_A + \sum_i \lambda_{i,A} r_{is,A} + \sum_{jk} \lambda_{jk,A}\left(\tfrac{3}{2} r_{js,A} r_{ks,A} - \tfrac{1}{2}\lambda_{jk} r_{s,A}^2\right) + \sum_{lmn} \lambda_{lmn,A}(\ldots)\right] \tag{13}$$

that shows clearly that points far away from atom A (and thus a small weight w_s) get a small redistributed charge.

Using the constraints, we obtain for the Lagrangian multipliers:

$$\begin{aligned} &Q_A - \sum_s q_{s,A} = 0 \\ &Q_A = \lambda_A \sum_s w_{s,A} + \sum_i \lambda_{i,A} \sum_s w_{s,A} r_{is,A} + \sum_{jk} \lambda_{jk,A} \sum_s w_{s,A}\left(\tfrac{3}{2} r_{js,A} r_{ks,A} - \tfrac{1}{2}\lambda_{jk} r_{s,A}^2\right) + \ldots \end{aligned} \tag{14}$$

$$\begin{aligned} &\mu_{t,A} - \sum_s q_{s,A} r_{ts,A} = 0 \\ &\mu_{t,A} = \lambda_A \sum_s w_{s,A} r_{ts,A} + \sum_i \lambda_{i,A} \sum_s w_{s,A} r_{ts,A} r_{is,A} + \sum_{jk} \lambda_{jk,A} \sum_s w_{s,A} r_{ts,A}\left(\tfrac{3}{2} r_{js,A} r_{ks,A} - \tfrac{1}{2}\lambda_{jk} r_{s,A}^2\right) + \ldots \end{aligned} \tag{15}$$

$$\Theta_{tu,A} - \sum_s q_{s,A}\left(\tfrac{3}{2} r_{ts,A} r_{us,A} - \tfrac{1}{2}\lambda_{tu} r_{s,A}^2\right) = 0$$

$$\Omega_{tu,A} = \lambda_A \sum_s w_{s,A}\left(\tfrac{3}{2} r_{ts,A} r_{us,A} - \tfrac{1}{2}\delta_{tu} r_{s,A}^2\right) + \sum_i \lambda_{i,A} \sum_s w_{s,A}\left(\tfrac{3}{2} r_{ts,A} r_{us,A} - \tfrac{1}{2}\delta_{tu} r_{s,A}^2\right) r_{is,A} + \sum_{jk} \lambda_{jk,A} \sum_s w_{s,A}\left(\tfrac{3}{2} r_{ts,A} r_{us,A} - \tfrac{1}{2}\delta_{tu} r_{s,A}^2\right)\left(\tfrac{3}{2} r_{js,A} r_{ks,A} - \tfrac{1}{2}\delta_{jk} r_{s,A}^2\right) + \ldots \quad (16)$$

and a similar equation for the octupole moments. This results for each atom in a set of linear equations (of size 4 when distributing up to dipole moment $[Q,\lambda_i]$, 10 up to quadrupoles $[Q,\lambda_i,\lambda_{jk}]$ and 20 up to octupoles $[Q,\lambda_i,\lambda_{jk},\lambda_{lmn}]$) for the Lagrangian multipliers, which are solved by a standard $Ax = b$ routine.

Finally, the values obtained for the Lagrange multipliers are used to get the *redistributed* atomic charges which when summed, result in the Multipole Derived Charges (up to some order X):

$$q_s^{MDC} = \sum_A q_{s,A} \quad (17)$$

In the following we shall refer to the charges as MDC-D charges if the multipoles are reconstructed up to the dipole moment, MDC-Q if up to quadrupole (recommended to be used) and MDC-O if up to the octupole moment.

Computational details

Molecule set

We investigated a set of 31 molecules[a] to obtain a good test of the quality of the method, that were taken predominantly from ref. [171], and extended with some that were of interest to us. Furthermore, as a second test set, we used all amino acid residues. Since we are, in *this* study, not interested in creating a generally applicable force field for amino acid residues for use in MD simulations of proteins, we took as a model for the amino acid residues a reduced conformation as they appear in proteins. I.e. replacing the NH_3^+ and COO^- groups by NH_2 and CHO. In fact, this means we cut off the backbone and replaced it by hydrogens. We are aware of the fact that this is not the standard model being used for getting amino acid residue charges, but we like to keep the model as simple as possible. Moreover, for a few amino acid residues, we also tested the so-called dipeptide model ($CH_3CONHCHRCONHCH_3$) for the side chain *R*, which resulted in virtually the same charges in the side-chain.

For all molecules, we first optimized the geometries, then we performed single point energy calculations to get the molecular properties (energies, multipoles). All calculations were carried out within the Density Functional Theory framework[1] with the ADF program[118,176-178] using the Becke-Perdew exchange-correlation[120,121] potential in the TZ2P (*V* in ADF terms) basis set. In the following we shall make a distinction between the *fitted* and *"exact"* molecular multipole moments: the former refer to the values obtained from the *fitted* density (or from the MDC charges since they represent them exactly), while the latter refers to the values from the *"exact"* density.

[a] The set of 31 molecules consisted of: benzene, ethylene, methylcyanide, methyllithium, methanol, methanol anion, methanethiol, methanethiol anion, methane, chlorine, carbonmonoxide, carbondioxide, carbondisulfide, fluorine, hydrogen, formaldehyde, hydrogenchloride, formamide, formic acid anion, hydrogenfluoride, hexafluorobenzene, hydrogen fluoride, lithiumhydride, nitrogen, ammonia, oxygen, dimethylether, hydroxyl anion, phosphine, thiophene, water, water dimer

Exponential prefactor

This left us the task to find a value for the exponential prefactor α, which we want to make as large as possible to keep the multipoles local. On the other hand, if we make it too big, the weight function will approach a delta function, leading to hardly any freedom to distribute the charges over the other atoms and numerical instabilities. We took as "optimal" value the largest value where the errors in the represented multipoles (due to machine precision and numerical accuracies) of the amino acid residues set were smaller than the required accuracy of the numerical integration in the calculations (10^{-6} in these cases), which turned out to be 3.0.

Buckingham convention

In the literature, several conventions are being used for the multipole moments. In this study, we use the most commonly used Buckingham convention[179-181]. This convention has the following expressions for the multipole moments:

$$\begin{aligned} \theta_i &= \int \rho r_i d\tau \\ \theta_{jk} &= \int \rho \left(\tfrac{3}{2} r_j r_k - \tfrac{1}{2}\delta_{jk} r^2\right) d\tau \\ \theta_{lmn} &= \int \rho \left(\tfrac{5}{2} r_l r_m r_n - \tfrac{1}{2}\delta_{lm} r_n r^2 - \tfrac{1}{2}\delta_{mn} r_l r^2 - \tfrac{1}{2}\delta_{nl} r_m r^2\right) d\tau \\ &\ldots \end{aligned} \tag{18}$$

In this convention, the electrostatic potential is obtained as follows:

$$V_C^A(r_s) = \frac{Q_A}{r_{sA}} + \sum_i \frac{\theta_i r_{sA,i}}{r_{sA}^3} + \frac{1}{2}\sum_{jk} \frac{\theta_{jk} r_{sA,j} r_{sA,k}}{r_{sA}^5} + \frac{1}{6}\sum_{lmn} \frac{\theta_{lmn} r_{sA,l} r_{sA,m} r_{sA,n}}{r_{sA}^7} + \ldots \tag{19}$$

Inserting the expressions for the Z_{lm}'s in the multipole expansion (eqn. (8)), and rewriting them into Cartesian components, the relation to the Buckingham multipoles can easily be derived.

All *molecular* multipole moments reported here have been obtained relative to the center of mass of the molecule.

Point charges in non-atomic (dummy) positions

For small systems, a proper charge distribution can not be represented by assigning charges only to atoms[1]. This is most easily seen for a homonuclear diatomic molecule like hydrogen. When using charges on the two atoms only, all methods should result in charges of exactly zero, because of symmetry. Adding a *third* (dummy) point, for instance in the center, should then result in a much better description of the charge distribution within the molecule (see also Allen/Tildesley[160,161], who use nitrogen as an example and use 5 points with rather large charges). In practice, adding dummy points is only necessary for small molecules, and in all cases adding one point (which is placed in bond midway points in this section) already suffices to reproduce the multipole moments up to the quadrupole moment.

ADF implementation

The method described in this section is generally applicable, however, we use Density Functional Theory[1] as incorporated in the ADF program[118,176-178] and take advantage of the way the electrostatic potential is being calculated there. The program uses a numerical integration scheme, by employing a

grid around the atoms to do the integration[176]. It uses a basis set of Slater type orbitals[117]:

$$\chi_u(r,\theta,\phi) = r^{n_u-1} e^{-\zeta_u r} Z_{lm}(\theta,\phi) \tag{20}$$

centered on each atom, with an auxiliary set of (Slater-type) fitfunctions (called *fit set*) to approximate the density ρ by expansion in these one-center functions f_i with coefficients a_i's:

$$\rho(\mathbf{r}) = \sum_{\mu\nu} P_{\mu\nu}\chi_\mu\chi_\nu \approx \sum_i a_i f_i(\mathbf{r}) \tag{21}$$

The atomic multipoles are obtained from the coefficients:

$$M_{lm}^A = \sum_{i\in A} a_i \int f_i(\mathbf{r}_2)\mathbf{r}_2^{\,l} Z_{lm}(\mathbf{r}_2)\, d\tau_2 \tag{22}$$

where the a-coefficients are obtained from a least squares minimization of the difference Ω_{AB} between the *"exact"* and *fitted* density:

$$\Omega_{AB} = \int \left|\rho_{AB} - \tilde{\rho}_{AB}\right|^2 d\tau \tag{23}$$

The electrostatic potential is now obtained as[118]:

$$V_C(r_s) = \sum_A V_C^A(r_s) = \sum_A \sum_{i\in A} a_i \int \frac{f_i(r)}{|r_s - r|} dr \tag{24}$$

or, by using the expansion of $|r_s - r|$ in spherical harmonics and using the exponential form of the STO fit functions (with the exponent ζ_i and the principal quantum number n_i):

$$\begin{aligned} V_C^A(r_s) &= \sum_l \sum_{m=-l}^{m=l} \frac{4\pi}{2l+1} Z_{lm}(r_{sA}) I_{lm}^A(r_{sA}) \\ I_{lm}^A(r_{sA}) &= \sum_{i\in A} \delta(l,l_i)\delta(m,m_i) a_i I(n_i,l_i,\zeta_i,r_{sA}) \end{aligned} \tag{25}$$

The function I is obtained from incomplete Gamma functions, and can be written as the sum of a multipolar and exponentially decaying part[118]:

$$I(n_i,l_i,\zeta_i,r_{sA}) = \frac{1}{(r_{sA})^{l_i+1}} \frac{(n_i+l_i+1)}{(\zeta_i)^{n_i+l_i+2}} + e^{-\zeta_i r_{sA}} J(n_i,l_i,\zeta_i,r_{sA}) \tag{26}$$

The function J consists of a power series in r_{sA}, with n_i as highest power, and serves as a screening for the short-range behavior of the multipole expansion. Since the expansion with the screening is correct inside the molecule, outside the molecule, where the screening is absent (due to the short range of action), the expansion is also valid and thus gives the correct potential.

A lot of effort has been put in constructing appropriate *fit sets*, and the current standard sets are very well qualified to reproduce electrostatic potentials inside the molecule accurately. Usually, the set of fit functions is larger than the set of basis functions, and sufficiently large to give a good description

of the electrostatic potential within the molecule. However, comparing the *fitted* molecular moments with the *"exact"* values from the basis functions, we found some differences, which were due to fit set incompleteness. To remedy this effect, we constructed a new set of fit functions by adding p and d-functions for the atoms involved in our test set of molecules, except for hydrogen that was not altered at all. With this new set, the differences between the *"exact"* and *fitted* molecular multipole moments were much smaller (see Table 3.1.1 for values averaged over the set of 31 molecules). From this Table, it can be seen that while the new fit set gives better *molecular* multipole moments, there is hardly any influence on the total energy.

TABLE 3.1.1. COMPARISON OF *STANDARD* AND *NEW* FIT SET. ABSOLUTE DIFFERENCES BETWEEN *"EXACT"* AND *FITTED* MOLECULAR MULTIPOLE MOMENTS, IN TOTAL ENERGY AND FIT TEST DIFFERENCE, ALL AVERAGED OVER THE SET OF 31 MOLECULES (A.U.)

	standard		*new*	
	average	maximum	average	maximum
Q_{tot}	0.000	0.000	0.000	0.000
Ω_x	0.004	0.039	0.004	0.060
Ω_y	0.002	0.023	0.001	0.014
Ω_z	0.038	0.303	0.026	0.258
Ω_{xx}	0.490	3.640	0.105	0.267
Ω_{xy}	0.047	0.807	0.035	0.413
Ω_{xz}	0.000	0.000	0.000	0.000
Ω_{yy}	0.373	1.788	0.102	0.251
Ω_{yz}	0.000	0.000	0.000	0.000
Ω_{zz}	0.685	3.577	0.139	0.502
Fit test	$5.05 \cdot 10^{-5}$		$3.95 \cdot 10^{-5}$	
Energy difference	0.029 kcal/mol			

Extensions of the implementation

The standard *fit* sets in the ADF program have been constructed in order to minimize the difference between the *"exact"* and *fitted* density, thereby obtaining from both accurate electrostatic potentials in the system under study. However, no specific constraints were put on the functions to conserve the molecular multipole moments. Moreover, in the standard versions of ADF, the expectation value of only the dipole moment (from the *"exact"* density) is being calculated. We implemented the calculation of the expectation value of the quadrupole moment tensor, in order to check the represented quadrupole moments from both the *"exact"* and *fitted* densities.

In Table 3.1.1, we give the absolute differences between the *molecular* multipole moments from the *"exact"* density on one hand, the *fitted* density on the other, averaged over the set of 31 molecules. We report the values for both the standard fit set (TZ2P basis set, V in ADF terminology) as well as a new one, where we added p- and d-functions to the fit set for the first row-atoms. We also give the values of the fit test for both *fit* sets, as well as the averaged absolute energy differences between the two *fit* sets. These numbers give a clear indication that there is hardly any influence on the total energy and *"exact"/fitted* density difference. There is however a clear improvement of the represented multipole moments, with an average absolute deviation of 0.10 au for the quadrupole moments.

TABLE 3.1.2 COMPARISON OF MOLECULAR MULTIPOLE MOMENTS (A.U.)

Multipole	*DFT Exact*	*DFT Fitted*	*Experimental*
benzene			
$\Theta_{xx} = \Theta_{yy}$	2.78	2.79	3.23
Θ_{zz}	-5.56	-5.59	-6.47 ± 0.37
carbonmonoxide			
Θ_{z}	-0.07	-0.09	-0.04
$\Theta_{xx} = \Theta_{yy}$	0.76	0.74	
Θ_{zz}	-1.52	-1.48	-1.86 ± 0.22
carbondisulfide			
$\Theta_{xx} = \Theta_{yy}$	-1.35	-1.58	
Θ_{zz}	2.71	3.15	3.17 ± 0.22
fluorine			
$\Theta_{xx} = \Theta_{yy}$	-0.38	-0.33	
Θ_{zz}	0.76	0.66	0.56
hydrogen			
$\Theta_{xx} = \Theta_{yy}$	-0.25	-0.31	-0.24
Θ_{zz}	0.51	0.61	0.47 ± 0.03
hexafluorobenzene			
$\Theta_{xx} = \Theta_{yy}$	-2.99	-2.78	-3.53
Θ_{zz}	5.99	5.56	7.07 ± 0.37
hydrogenfluoride			
Θ_{z}	-0.69	-0.69	-0.72
$\Theta_{xx} = \Theta_{yy}$	-0.87	-0.95	
Θ_{zz}	1.74	1.91	1.75 ± 0.02
nitrogen			
$\Theta_{xx} = \Theta_{yy}$	0.561	0.490	0.76
Θ_{zz}	-1.122	-0.979	-1.52
oxygen			
$\Theta_{xx} = \Theta_{yy}$	0.071	0.065	0.20
Θ_{zz}	-0.142	-0.131	-0.39
water			
Θ_{z}	-0.71	-0.73	-0.73
Θ_{xx}	-1.82	-1.93	-1.86 ± 0.01
Θ_{yy}	1.91	2.01	1.96 ± 0.01
Θ_{zz}	-0.09	-0.08	-0.10 ± 0.02

Results

Multipole moments

Although this section is strictly speaking not concerned with the accuracy of calculating molecular multipole moments, at least it should be established that Density Functional Theory is able to provide reasonable values for them. As the atomic charges are constructed to represent these molecular moments, in order to get reliable charges the molecular multipoles should be correct. In Table 3.1.2, a few calculated molecular multipole moments are given with the corresponding experimental values. The calculated values are given both resulting from the *"exact"* and *fitted* density to be able to compare the difference between them also, and are obtained in the TZ2P-newfit basis set with the Becke[120]-Perdew[121] exchange-correlation potential.

For benzene, the first non-vanishing multipole moment is (due to symmetry) the quadrupole moment. The values for the axes lying in the benzene plane are equal, and are minus half times the perpendicular value (also due to symmetry). The experimental value of −6.47 a.u. for the perpendicular value Θ_{zz} has been obtained by induced-birefringence experiments, and is reasonably well described by the DFT values of −5.55 (*"exact"* density) and −5.59 a.u. (*fitted* density), if the newfit set is used. Benzene is a good example why the standard fit sets are not always adequate for use with the new charge analysis presented here. In the standard fit set, the Θ_{zz} value from the *"exact"* density doesn't change much compared to the newfit set (−5.60 a.u.); the value from the *fitted* density on the other hand differs considerably (−9.19 a.u.). The dipole moment of carbon monoxide has in the past served as a test case for assessing the quality of wavefunction based methods, as it was shown that for instance Hartree-Fock predicts a dipole moment that points in the wrong direction. I.e., the negative end of the molecule was predicted on the side of the oxygen, while in reality it should be on the side of the carbon. In fact, the methods where electron correlation is included explicitly (like for instance MP2, MRSDCI or CASSCF) give a proper qualitative picture. As can be seen from Table 3.1.2, the same is true for the Becke-Perdew xc-potential. In fact, in a recent paper by de Proft et al.[182], several other DFT potentials were tested and all were shown to give a proper description as well. The same agreement between calculated and experimental values is found for the quadrupole moments of carbondisulfide, fluorine, hydrogen, hexafluorobenzene, nitrogen and oxygen. Hexafluorobenzene is another striking example of the need for improved fit sets, as the value for its Θ_{zz} value from the *fitted* density of the standard TZ2P fit set is 2.81 a.u., while the value from the *"exact"* density is 6.08 a.u. in this standard basis set. Although the difference between the *"exact"* and *fitted* value is a bit larger than for benzene (0.43 a.u., see Table 3.1.2) with the new fit set, it is still a considerable improvement. The calculated values for the dipole and quadrupole moments of hydrogenfluoride and water are again in good agreement with the experimental values.

Charges

The MDC-q charges of some molecules out of the set of 31 are given in Table 3.1.3. For a few molecules, point charges in non-atomic (dummy) positions have been used as described earlier in this section. In all cases, the molecular multipole moments up to the quadrupole moment have been represented (*by construction*) by these charges. Benzene and hexafluorobenzene show the predicted trend; in benzene, the electrons are more likely to be found on the carbons, while in hexafluorobenzene, they are attracted by the more electronegative fluoride atoms, just as would be expected from the change in sign of the Θ_{zz} value when going from benzene to hexafluorobenzene. There are also some interesting features when comparing the diatomic molecules. Where in the case of chlorine, fluorine and hydrogen, the charge is mainly found in-between the two atoms (as indicated by the charge on the dummy point XX), for carbon monoxide, nitrogen and oxygen, it is found mainly on the atoms. This is of course fully consistent with the normal picture of chemical bonding for the first part (where the charge is mainly found in-between the atoms), while the cases of CO, N_2 and O_2 can be explained by the lone pair electrons that are found on the atoms. Although from the MDC-q charges (including a dummy point) for carbon monoxide, it seems that the more negative end is found near the oxygen, the MDC-d charges do give the correct representation. And as the molecular multipole moments are exactly reproduced up to the quadrupole by the MDC-q charges, also they give a proper description for the dipole moment.

For a molecule with a non-singlet ground state (like for instance oxygen), the charge density can be obtained for both spins separately. E.g., for oxygen with two alpha electrons in excess (triplet state), one obtains a different charge density for the alpha and beta electrons. Resulting from these are also

two sets of *atomic* multipoles; one for the alpha electrons and one for the beta electrons. From these we can extract *alpha* and *beta* Multipole Derived Charges, which when summed result in the atomic charge again. However, the difference between the two densities is the spin density, and as such, the difference between the two sets of charges, the spin density charges. For oxygen, the total charge on the atoms is small (-0.049), but the spin density charge large (-1.007), which is fully consistent with the triplet state with one alpha electron located on one atom, and one on the other.

TABLE 3.1.3 MDC-Q CHARGES FOR SOME OF THE MOLECULES OF THE SET OF 31

Molecule	Charges	Molecule	Charges
benzene		*hydrogen*	
q_C	-0.123	q_H	0.610
q_H	0.123	q_{XX}	-1.220
ethylene		*hexafluorobenzene*	
q_C	-0.250	q_C	0.094
q_H	0.125	q_F	-0.094
methylcyanide		*nitrogen*	
$q_{C\text{-}Me}$	0.692	q_N	-0.452
q_H	-0.157	q_{XX}	0.904
q_C	0.203	*ammonia*	
q_N	-0.425	q_N	-0.444
chlorine		q_H	0.148
q_{Cl}	0.367	*oxygen*	
q_{XX}	-0.734	q_O	-0.049
carbondioxide		q_{XX}	0.098
q_C	0.574	*water*	
q_O	-0.287	q_O	0.458
carbonmonoxide		q_H	0.892
q_C	-0.537	q_{XX}	-1.121
q_O	-0.852	*water dimer*	
q_{XX}	1.389	$q_{H\text{-}donor}$	0.902
carbondisulfide		$q_{XX\text{-}donor}$	-1.162
q_C	-0.351	$q_{O\text{-}donor}$	0.264
q_S	0.185	$q_{H\text{-}donated}$	0.699
fluorine		$q_{XX\text{-}donated}$	-0.748
q_F	0.186	$q_{O\text{-}acceptor}$	0.560
q_{XX}	-0.371	$q_{H\text{-}acceptor}$	0.930
		$q_{XX\text{-}acceptor}$	-1.188

For carbondioxide and carbondisulfide, the sign of the quadrupole moment is reversed; i.e., whereas Θ_{zz} is negative for carbondioxide, for carbondisulfide it is positive. This is reflected by the charge of the carbon atom, which is positive for the former and negative for the latter. The larger electronegativity of oxygen over sulphur plays a decisive role also here. The molecular dipole and quadrupole moment of water can not be represented exactly by charges on the atoms only; putting charges on dummy positions on both bond midway-points is however sufficient to do this. This results in rather large charges in the dummy positions, which are again indicative for the chemical bonding taking place. For the *hydrogen bonded* water dimer, depicted in Figure 3.1.1 with the donor on the left

and the acceptor on the right, the point charges in the dummy positions were also included. This enables a comparison with the values for water itself. Not surprisingly, the charges for the hydrogen (and corresponding dummy point charges) not involved in the hydrogen bond do not change significantly. The charge on the hydrogen involved in the hydrogen bond (the so-called *donated* hydrogen) changes dramatically. Whereas the total of the hydrogen and dummy charge is around −0.23 in water, for the *donated* hydrogen this drops down to −0.05; the difference is mainly flowing to the oxygen of the donor molecule, whose positive charge drops down by almost the same amount.

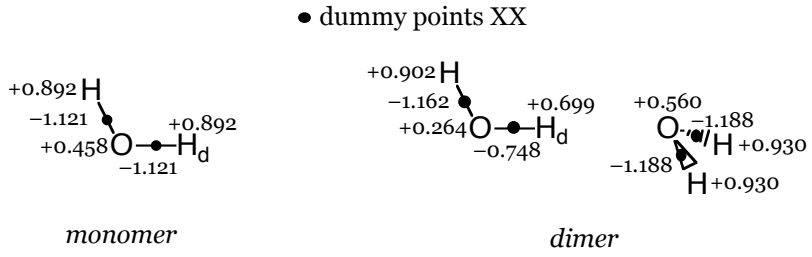


FIGURE 3.1.1 WATER MONOMER AND DIMER WITH MDC-Q CHARGES

The charges for the amino acid residue models *in vacuo* we have considered are given in Table 3.1.4 for the amino acids that are found in the active site of wildtype azurin, to be able to compare them with the values found within the active site (Section 6.1).

TABLE 3.1.4 AMINO ACID RESIDUE CHARGES FOR CYS, GLY, HIS AND MET

Atom	Charge	Atom	Charge	Atom	Charge	Atom	Charge
cysteine		*glycine*		*histidine*		*methionine*	
N	-1.028	N	-0.798	N	-0.882	N	-0.882
H	0.299	H	0.237	H	0.287	H	0.332
H	0.350	H	0.232	H	0.291	H	0.280
CA	0.461	CA	1.076	CA	0.505	CA	0.361
HA	-0.139	HA2	-0.249	HA	-0.087	HA	-0.103
CB	0.398	HA3	-0.247	CB	0.440	CB	0.496
HB2	-0.106	C	0.316	HB2	-0.109	HB2	-0.113
HB3	-0.117	O	-0.517	HB2	-0.178	HB3	-0.203
SG	-0.886	H	-0.051	CG	0.062	CG	0.176
C	0.319			ND1	-0.613	HG2	-0.033
O	-0.536			CE1	0.382	HG3	-0.057
H	-0.014			HE1	0.081	SD	-0.298
				NE2	-0.537	CE	-0.142
				HE2	0.381	HE1	0.079
				CD2	-0.035	HE2	0.104
				HD2	0.132	HE3	0.089
				C	0.390	C	0.367
				O	-0.471	O	-0.434
				H	-0.039	H	-0.021

Conclusions

The charge analysis presented here gives an accurate description of the charge distribution when obtained with an appropriate set of fit functions as we have introduced in this paper. There are a few advantages of the method: the atomic multipoles that are reconstructed exactly by our Multipole Derived Charges, give by construction the best fit to the electrostatic potential around the molecule. Furthermore, with the appropriate set of fit functions the molecular multipole moments from the fitted density can be made arbitrarily close to the ones from the "exact" density. Our Multipole Derived Charges reproduce the fitted multipoles exactly, except in few cases (like diatomics) when there are not enough degrees of freedom to reproduce all multipoles. Usually it is sufficient to add one extra point, e.g. in the bond middle point of a diatomic molecule.

The molecular multipole moments, both from the fittted and the "exact" density, have a small deviation from experimental values. Since the MDC charges reproduce the fitted values exactly, the MDC charges give an accurate representation of the charge distribution in the molecule and result in a good electrostatic potential.

3.2 *IntraFF force constants*

Obtaining reliable force constants for use in classical Molecular Dynamics simulations

Over the last decades, many investigations have focused on proteins containing metal atoms, both from an experimental and theoretical point of view (see Section 1.2). The theoretical studies comprise mainly classical Molecular Dynamics[160] (MD) studies, but in the last years it has become possible to treat (reasonable models of) the active sites of these proteins accurately by quantum chemical methods[1]. However, at the present time, it is still not possible to treat the complete protein with quantum chemical (QC) methods nor to perform reliable MD simulations (of at least 1 ns) with (part of) the active site treated at the QC level. For the time being, one therefore has to treat the active site completely classically. This raises a problem, because the interactions of metal atoms are difficult to generalize in terms of a simplified classical force field[1]. We resolve this issue by performing a QC calculation prior to the MD simulation to obtain force field parameters that can be used subsequently in the simulation and which are typical for the particular system under study. The necessary parameters can be divided into two groups: *bonded* and *non-bonded*. The latter consist of Lennard-Jones parameters (which have already been parameterized in the GROMOS96 force field[127]) and atomic fractional charges. The charges can be obtained accurately with Density Functional Theory calculations by using the Multipole Derived Charge analysis[183]. We are therefore left with obtaining parameters for *bonded* interactions.

The most commonly used way to treat these bonded interactions classically is by using a harmonic potential for bonds and bond angles:

$$U(C) = \tfrac{1}{2} K_C \left(C - C_{eq}\right)^2 \tag{1}$$

where C_{eq} is the equilibrium value of the coordinate (bond/angle), and K_C a force constant typical for that particular bond/angle. A small force constant implies great flexibility in the coordinate while a large value is found for stiff bonds/angles. For dihedral angles ϕ one normally takes a periodic function:

$$\begin{aligned} U(\phi) &= K_\phi \left(1 + \cos\left(n\phi - \phi_{shift}\right)\right) = K_\phi \left(1 + \cos\left(n\left(\phi - \phi_{eq}\right) - \pi\right)\right) \\ \phi_{shift} &= n\phi_{eq} + \pi \end{aligned} \tag{2}$$

where n stands for the periodicity of the dihedral angle. Note that GROMOS96[127] uses a quartic potential (see **Appendix 3.2**) for bonds, which was introduced only to avoid the square root operation in the calculation. The deviations from the harmonic potential however show up only at high energies (i.e. at large deviations from the equilibrium position) that are normally never reached in simulations. The IntraFF method can also be used to obtain the force constant for this quartic potential.

Seminario proposed a method (FUERZA)[184] to get the force constants (FC's) directly from the Hessian matrix (a matrix of second derivatives of the energy with respect to atomic coordinates) as obtained in QC calculations. The method is invariant for the choice of internal coordinates, but the particular formulas he used for getting the force constants are not related directly to the classical

energy expressions. They result for instance in low values for bending constants, which is corrected for by adding extra *nonphysical* bonds. In his definition, two atoms are bonded if the energy increases when one of the two is moved in any direction, no matter how far apart the atoms are. For instance, the hydrogen atoms in water are bonded with a force constant of 31 kcal/mol/Å^2. Although it is true that these hydrogens are "bonded", the interaction between them has already been accounted for by the angle (bending) interaction.

In this study we report a new method (IntraFF) for obtaining force constants from QC calculations, that is based on ideas by Seminario[184], but with formulas taken from the classical energy expressions as used in the force fields, in which the total energy is a sum of contributions from separate internal coordinates (like bonds, angles and dihedrals). Each of these contributions consists of a force constant K_C and a F_C part:

$$U_{total} = \sum_C U_C = \sum_C K_C F_C \tag{3}$$

The F_C part depends on the coordinates of **only** those atoms that are involved in the internal coordinate *C*. Likewise, the Hessian (*3Nx3N* matrix for a system of *N* atoms containing the second derivatives of the energy with respect to the atomic Cartesian coordinates) is also a sum of contributions from the separate internal coordinates. The force constants K_C are obtained from the quantum chemical Hessian; they could be obtained by fitting the Hessian from the force field (FF) to the quantum chemical Hessian, but by doing so one already has to make a choice for the coordinates to be included in the force field, and the force constant values will depend on this choice. Instead, in the IntraFF method a direction vector is chosen for each internal coordinate *C*, which is characteristic for the coordinate; if the atoms involved in a coordinate are displaced by an infinitesimal amount along the direction vector of the coordinate, it will lead to a force working on the atoms, which is determined by the relevant part of the (QC or FF) Hessian. By minimizing the difference between the QC and FF infinitesimal forces, an expression is obtained in which the force constants are uniquely determined (see eq. (14)).

Therefore, just like the FUERZA scheme of Seminario[184], our method is invariant for the choice of coordinates used. The equations for harmonic potentials applied to molecules in their equilibrium structure will be derived first. Then, we will discuss the specific details for the direction vectors needed for these equations, and the harmonic equations will be extended for molecules in a non-equilibrium structure. Subsequently, the equations for an anharmonic potential are presented, followed by the a normal mode analysis part and finally the results for some molecules.

A final subject that should be addressed here concerns the nomenclature of the different coordinate systems that are in use. Although the term *internal coordinates* has been used for many different kinds of coordinate sets, we use it for the *molecular* internal coordinates such as bonds, angles and dihedral angles. The coordinate system consisting of the normal mode vectors will be referred to as *normal mode coordinates*. Together with the *Cartesian coordinates* these sets of coordinates are complementary and sufficient.

Harmonic potentials for molecules in equilibrium geometries

First, we will have to make the connection between the quantum chemical description and the classical force field description. This can be achieved if we expand the energy U to second order in a Taylor series about the atomic coordinates r. For this we need a direction vector $d = \delta r$ which has a special

meaning in our analysis (more details will be given later on); it will be shown that it should be representative for a certain *internal* coordinate. For the moment however, we will keep the discussion general and let the vector be arbitrary. We then get the following expression for the Taylor series (up to second order):

$$U(r+d) = U(r) + d^T G(r) + \tfrac{1}{2} d^T H(r) d \tag{4}$$

where $G(r)$ stands for the gradient and $H(r)$ for the Hessian, with for atoms i, j the elements:

$$\begin{aligned} G_{i,p} &= \frac{\partial U(r)}{\partial r_{i,p}} \\ H_{i,p;j,q} &= \frac{\partial^2 U(r)}{\partial r_{i,p} \partial r_{j,q}} \\ r_{i,p} &\quad \{x_i, y_i, z_i\} \end{aligned} \tag{5}$$

By definition, an equilibrium geometry is characterized by a zero gradient. Thus, the Taylor expansion leaves us with an expression for the change in energy that contains only a quadratic term (just like the classical harmonic expression [eq. (1)]):

$$U\left(r_{eq} + d\right) - U\left(r_{eq}\right) = \tfrac{1}{2} d^T H\left(r_{eq}\right) d \tag{6}$$

In principle, we could try and fit the Hessian matrix of the classical force field to the quantum chemical Hessian matrix. However, our method would then not be invariant to the choice of internal coordinates any longer, since we would have to choose which internal coordinates are used in the fit.

It would be more convenient to obtain the force constant for a given *internal* coordinate as if it were an independent parameter, like it is used also in most of the popular force fields. This can be achieved by looking at the change in gradient if a certain *internal* coordinate is changed by an infinitesimal amount, as proposed by Seminario[184]. For this, we expand not the energy but the gradient in a Taylor series, and obtain a formula for the gradient change δG (called *infinitesimal* gradient) that is linear in both the Hessian and the direction vector:

$$G(r+d) - G(r) = \delta G = H(r) d_C \tag{7}$$

This expression is valid for both the classical (FF) and the quantum chemical (QC) Hessian, and results in the classical (δG_{CL}) and quantum chemical (δG_{QC}) *infinitesimal* gradients. At this point, we have used the direction vector d_C with its special meaning as will be discussed later on, since it should be representative for the internal coordinate we are interested in. We will delay the discussion about the form of this direction vector until later, since its specific details are not important for formulating the analysis, as long as the vector is representative for a certain *internal* coordinate. Now, we want to find a value for the force constant such that the difference between the FF and the QC *infinitesimal* gradient is minimized. This is achieved by using a least squares fit of the difference between the *infinitesimal* gradients:

$$\Delta_G = \tfrac{1}{2} \left(\delta G_{CL} - \delta G_{QC}\right)^T \left(\delta G_{CL} - \delta G_{QC}\right) \tag{8}$$

The best value for the force constant K_C is then obtained if the derivative of this least squares function with respect to the force constant is zero:

$$\frac{\partial \Delta_G}{\partial K_C} = 0 \rightarrow \left(\delta G_{CL} - \delta G_{QC}\right)^T \frac{\partial\left[\delta G_{CL} - \delta G_{QC}\right]}{\partial K_C} = 0 \rightarrow \left(\delta G_{CL} - \delta G_{QC}\right)^T \frac{\partial\left[\delta G_{CL}\right]}{\partial K_C} = 0 \tag{9}$$

The last part results from the fact that only the classical Hessian (and therefore the classical *infinitesimal* gradient) depends on the force constants. The derivative of this *infinitesimal* classical gradient with respect to K_C is obtained from the derivative of the classical Hessian with respect to the force constant. I.e., substituting the expression for the infinitesimal gradient (eq. (7)) into the last part of eq. (9), and taking the direction vector d_C outside the derivative expression, we obtain:

$$\left(\delta G_{CL} - \delta G_{QC}\right)^T \frac{\partial\left[H_{CL}(r) d_C\right]}{\partial K_C} = 0 \rightarrow \left(\delta G_{CL} - \delta G_{QC}\right)^T \left[\frac{\partial H_{CL}(r)}{\partial K_C} d_C\right] = 0 \tag{10}$$

Now, we have to establish how the classical Hessian depends on the force constant K_C. In the harmonic approximation, the classical Hessian has the following form:

$$H^{CL}_{r_{i,p};r_{j,q}} = \sum_C K_C \left[\left(C - C_{eq}\right) \frac{\partial^2 C}{\partial r_{i,p} \partial r_{j,q}} + \frac{\partial C}{\partial r_{i,p}} \frac{\partial C}{\partial r_{j,q}}\right] = \sum_C K_C H^*_C \qquad r_{i,p} \quad \{x_i, y_i, z_i\} \tag{11}$$

It is a sum over products of the force constants K_C of the internal coordinates times a matrix that depends only on the coordinates of the atoms involved in that internal coordinate. The derivative of the total Hessian with respect to the force constant K_C of one coordinate is then **independent of the other coordinates**, and independent of K_C:

$$\frac{\partial H_{CL}}{\partial K_C} = H^*_C \tag{12}$$

After putting this result into eq. (10), the only term in it that still depends on the value of K_C is the *infinitesimal* gradient δG_{CL}. We make it more explicit by combining the outcome of eq. (11) with eq. (7):

$$\delta G_{CL} = H_{CL} d_C = \left\{K_C H^*_C\right\} d_C = K_C \left(H^*_C d_C\right) = K_C V_C \tag{13}$$

It results in a multiplication of the scalar K_C with a vector V_C, which can be combined with eq. (10) to finally get an expression for the force constant K_C:

$$\begin{aligned} &\left(\delta G_{CL} - \delta G_{QC}\right)^T \left[H^*_C d_C\right] = 0 \rightarrow \delta G_{CL}{}^T \left[H^*_C d_C\right] = \delta G_{QC}{}^T \left[H^*_C d_C\right] \rightarrow \\ &K_C V_C{}^T \left[H^*_C d_C\right] = \delta G_{QC}{}^T \left[H^*_C d_C\right] \rightarrow K_C = \frac{\delta G_{QC}{}^T \left[H^*_C d_C\right]}{V_C{}^T \left[H^*_C d_C\right]} = \frac{\left[H^*_{QC} d_C\right]^T \left[H^*_C d_C\right]}{\left[H^*_C d_C\right]^T \left[H^*_C d_C\right]} \end{aligned} \tag{14}$$

Although we have used the harmonic potential as an example here, this expression can in fact be used for any potential for which the Hessian can be written as a product as in eq. (11). Therefore, eq. (14) is also valid for the quartic GROMOS96[127] potential as well as the cosine function that is normally used for proper dihedrals (see **Appendix 3.2** for the expressions of the classical Hessian of the potentials described in this paper).

Direction vectors

Before going on to the expressions for getting force constants of molecules in a non-equilibrium position, one issue should be addressed. It involves the direction vectors (of unit length; in principle we can take a vector of arbitrary length, since it is involved in both the numerator and denominator of eq. (14), but for simplicity we take this direction vector as a unit vector) belonging to the *internal* coordinates for which we want to obtain force constants and/or equilibrium values. As shown in the harmonic potential part, the difference between the classical and quantum chemical *infinitesimal* gradients in the direction of this vector is minimized (see eq. (8)); but no specific description has been given yet what this vector looks like. The only thing stated so far, is that the direction vector d_C should be representative for the *internal* coordinate we are interested in. In this part, we shall deal with this issue.

As the interactions for a given *internal* coordinate are expressed in terms of a scalar value, the issue does not arise when using the coordinate in a classical force field. Hence, no clearcut formulations are readily available for it. We will have to make a choice for the direction vector based on the positions of the atoms involved in the *internal* coordinate. The most simple approach is to take the vector which will induce the largest increase in the value of the *internal* coordinate, which is given by the gradient of the coordinate. For example, for a bond between two (homonuclear) atoms, the direction vector is given in Figure 3.2.1 as δr:

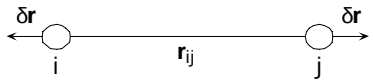


FIGURE 3.2.1. SCHEMATIC REPRESENTATION OF BOND DIRECTION VECTOR

Only in the case of a homonuclear dimer is this the one and only choice. For instance, if we take a heteronuclear dimer, we have to take the different atomic masses into consideration by using some kind of weight factor. But in what way, is not yet clear. We could argue that since heavy atoms move less than light atoms, the latter should have a higher weight; based on the expression for the kinetic energy, it would lead to a weight that is equal to the inverse square root of the atomic mass. But by looking at the force, we could argue that the weight should be equal to the inverse of the mass. Moreover, when using the latter choice for the direction vector, neither the center of mass nor the moments of inertia of the molecule change. Therefore, this choice is the preferred one, leading to the direction vector of a given *internal* coordinate C as:

$$d_{C,i} = w_i \left(\quad C \right)_i \tag{15}$$

where the weight for atom i is given as:

$$w_i = \frac{1}{M_i} \tag{16}$$

In order to check that this is indeed the most appropriate choice for the direction vector, we have tested a few other choices; we looked at the gradients, vibrational frequencies and fit residues (eq. (8)) for a number of molecules. For "normal" bonds, i.e. not involving metal atoms, the choice given in eq. (15) gives indeed the best agreement between quantum chemical and classical properties (gradients, vibrational frequencies, fit residues). However, as will be shown in the **Results** part, when looking at bonds involving metal atoms, more appropriate direction vectors can be used where only the metal atom is allowed to move.

Harmonic potentials for molecules in non-equilibrium geometries

In the harmonic potential part, we discussed how to obtain harmonic force constants for molecules in their equilibrium positions. However, most of computational chemistry is being done on molecules that are not *exactly* in the equilibrium geometry due to numerical accuracy. It would be most helpful to be able to correct the force constants as well as the equilibrium values of the bond lengths/angles for this non-equilibrium character. Within the harmonic approximation this is easily achieved by first getting the force constant as if the molecule were in an equilibrium geometry. Since the second derivative of the harmonic potential is constant and equal to the force constant (in 1D; in 3D, the trace of the Hessian is constant), the value of it does not change if the molecule is put into a non-equilibrium position. The only change occurs for the gradient, which measures how far the molecule is outside its equilibrium position:

$$G\left(C_{now}\right) = K_C\left(C_{now} - C_{eq}\right) \tag{17}$$

Since the gradient is known from quantum chemistry, the real equilibrium value can be obtained from this expression since all other quantities are known:

$$C_{eq} = C_{now} - \frac{G\left(C_{now}\right)}{K_C} \tag{18}$$

In three-dimensional notation, this would result in the following expression:

$$C_{eq} = C_{now} - \frac{\mathbf{G}\left(r_{now}\right)^T \; C}{K_C\left(\; C^T \; C\right)^{1/2}} \tag{19}$$

Note that if the gradient is zero, the equilibrium value is simply the current one, like already mentioned earlier.

For a proper dihedral, we do not use a harmonic potential but a cosine function. The equilibrium value can then be obtained by following the same strategy, resulting in:

$$\phi_{eq} = \phi_{now} + \frac{1}{n}\arcsin\left(\frac{-1}{nK_\phi}\frac{G\left(r_{now}\right)^T \; \phi}{\left(\; \phi^T \; \phi\right)^{1/2}}\right) \tag{20}$$

Anharmonic Frost potential

Until now, we have concerned ourselves with harmonic potentials, which are valid only if one is reasonably close to the equilibrium geometry of the molecule (see Figure 3.2.2). However, sometimes one might want to use an anharmonic potential that is able to give a proper description over a larger bond value range, for instance in the case of the dissociation reaction between any atoms or molecules. The most widely known anharmonic potential is due to Morse[130]. It uses the equilibrium distance R_{eq}, dissociation energy D_e and anharmonicity constant β:

$$U_{Morse} = D_e\left(1 - e^{-\beta\left(R - R_{eq}\right)}\right)^2 \tag{21}$$

Disadvantageous for this equation is the fact that it is a three parameter equation, while in the strictest sense, we only have two variables to optimize the parameters (first and second derivative: gradient and Hessian).

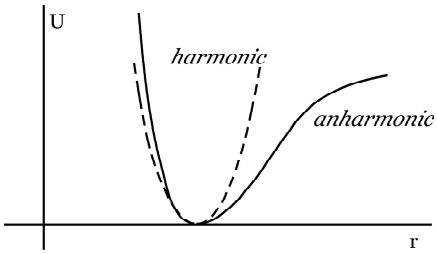


FIGURE 3.2.2. SCHEMATIC COMPARISON OF ANHARMONIC FROST AND HARMONIC POTENTIAL

Varshni[185] has examined several potentials with differing number of parameters, ranging from two to five. One of them is the two parameter Frost potential[131,132], which has been constructed on semi-theoretical grounds. It is split up in a nuclear and an electronic part, just like normally done in quantum chemistry:

$$U(R) = \frac{N_1 N_2}{R} + U_e(R) \tag{22}$$

N_i stands here for the nuclear charge of atom i, and U_e for the electronic part of the potential. Frost and Musulin then defined some theoretical criteria that should be satisfied by a given potential (they used also two less clear ones, which we will not show here):

1- the potential energy is the sum of two parts (i.e. nuclear and electronic)
2- U becomes infinite as R approaches zero
3- U_e is finite at $R = 0$
4- $U_e \to -N_1 N_2 / R$ for large R (due to choice of $U = 0$ at infinite R)
5- U must be capable of going through a minimum as R varies

A simple function for the electronic potential that satisfies these criteria, is given by:

$$U_e = -\frac{N_1 N_2}{R}\left(1 - e^{-aR}\right) - be^{-aR} \tag{23}$$

where a,b are adjustable parameters. They arrived at the particular form for this function by considering possible products of an exponential with powers of R, which was the most probable form because the overlap, Coulomb and exchange integrals of quantum-chemical theory are of this type when using Slater type orbitals. The sum of the electronic and nuclear potential energy is then given by:

$$U = e^{-aR}\left(\frac{N_1 N_2}{R} - b\right) \tag{24}$$

The function is plotted with parameters for the hydrogen molecule (from the original article[131]) in Figure 3.2.2. Examining several properties of 23 diatomic molecules, Varshni concluded that the Frost and Morse potential perform equally well[185].

The question remains however, how to obtain values for the a,b parameters. For that, we need not only the potential energy, but also the gradient and Hessian (in 1D):

$$\begin{aligned} U &= e^{-aR}\left(\frac{N_1 N_2}{R} - b\right) \\ G &= e^{-aR}\left(ab - \frac{aN_1 N_2}{R} - \frac{N_1 N_2}{R^2}\right) \\ H &= e^{-aR}\left(\frac{2N_1 N_2}{R^3} + \frac{2aN_1 N_2}{R^2} + \frac{a^2 N_1 N_2}{R} - a^2 b\right) \end{aligned} \tag{25}$$

For the equilibrium bond length the gradient is zero; therefore we have an expression for the b parameter:

$$ab - \frac{aN_1 N_2}{R_{eq}} - \frac{N_1 N_2}{R_{eq}^2} = 0 \rightarrow b = \frac{N_1 N_2}{R_{eq}} + \frac{N_1 N_2}{aR_{eq}^2} \tag{26}$$

We can use this in the equation for the Hessian, which results in an expression for it that depends only on a and R:

$$H = N_1 N_2 e^{-aR}\left(\frac{2}{R^3} + \frac{2a}{R^2} + \frac{a^2}{R} - \frac{a^2}{R_{eq}} - \frac{a}{R_{eq}^2}\right) = K' \tag{27}$$

However, we also have the value of the Hessian in this point (the force constant K within the harmonic approximation). Therefore, we can get the parameter a by minimizing the difference between the actual 2nd derivative K and the Frost 2nd derivative K':

$$\Delta_a = \tfrac{1}{2}\left(K - K'\right)^2 \tag{28}$$

Note, that in order to do this we have to choose a trial equilibrium bond length (which we will optimize later). From eq. (28) we get the value for the parameter a, which we can use in eq. (26) for the parameter b at a certain "equilibrium" bond length. With the a,b parameters, we obtain a value for the gradient within the Frost approximation. If it is equal to the actual quantum chemical gradient, the "trial" equilibrium bond length is correct, else we use the gradient difference to get a better guess of the equilibrium bond length, and repeat this procedure until we have found the correct one (within the Frost approximation).

With the optimized parameters a,b and the equilibrium bond length, the dissociation energy D_e is simply obtained by filling in these values in eq. (24). We then get:

$$D_e = -e^{-aR_{eq}}\left(\frac{N_1 N_2}{R_{eq}} - b\right) \tag{29}$$

Note that the minus sign has been added since the dissociation energy is normally defined as a positive quantity, i.e. the energy needed to dissociate a molecule and not as a negative interaction energy as in equation (24).

Normal mode analysis

An important check on the accuracy of the force field parameters from any method is the ability of them to reproduce the vibrational frequencies (be it from experiments or quantum chemical calculations). They can be obtained from the calculated Hessian in Cartesian coordinates, by transforming it into the mass-weighted Hessian matrix H^{mw}[186]:

$$H_{ij}^{mw} = \frac{H_{ij}}{\sqrt{M_i M_j}} \tag{30}$$

After diagonalization of this matrix, the eigenvalues ε_i can be transformed easily into the vibrational frequencies ν_i:

$$\nu_i = \frac{1}{2\pi}\sqrt{\varepsilon_i} \tag{31}$$

This procedure works for any Hessian matrix (both classical and quantum chemical), and therefore we can use it to check the reliability of the classical force field.

One important issue involved in this procedure is the coordinates involved in the normal mode analysis. For a molecule consisting of N atoms, there are $3N-6$ normal modes and vibrational frequencies ($3N-5$ for a linear molecule). So far, we have not been concerned about the number of classical *internal coordinates*, but it is obvious that they should be chosen such that the total number is equal to the number of normal modes. However, this is a choice of the person who wants to use the force constants for a classical simulation, and as such does not concern us here in the method development. Normally, the choice for which *internal coordinates* to use is easily made, like in the case of water. One would normally take two O-H bonds and one H-O-H angle, or alternatively three bonds: two O-H and one H-H. Moreover, it is even possible to combine several *internal coordinates* together in one *combined internal coordinate* as are standard in for instance AMBER95 with a few

dihedral angles. Although the total number of actual *internal coordinates* used is larger than the number of normal modes, this is not true when the *combined internal coordinates* are counted.

Computational details

The method described in this paper is generally applicable to any quantum chemical method able to supply a gradient and Hessian for a certain molecule in a given geometry. In this paper we will report results for some molecules where the quantum chemical data were obtained with Density Functional Theory (DFT)[1] calculations in the ADF program[117,187]. Although the results in this paper are used to test the method, and not the accuracy of DFT, we would like to obtain calculated values that are at least reasonable, to be able to compare the force field parameters obtained with standard force fields like AMBER[126]/GROMOS[127,176]. We therefore performed the DFT calculations in a TZP basis set using the LDA (Local Density Approximation)[1] exchange-correlation potential. In another study, we have investigated the accuracy of several exchange-correlation potentials in a number of basis sets for a set of 19 small molecules. The same set had been used previously by another group[188] to test the accuracy of several post-Hartree Fock methods in combination with different (correlation consistent) basis sets. This enabled us to compare our results not only to experimental data, but also to high-level *ab initio* calculations. We found that (for the TZP basis set we used) the accuracy of the LDA exchange-correlation potential is equal to the normally more accurate but also more CPU expensive Becke[120]-Perdew (BP86[121]) or Perdew-Wang (PW91[189]) exchange-correlation potentials: the average absolute deviations of the bond lengths from experimental data are respectively 1.14 (LDA), 1.24 (BP86) and 1.09 (PW91) pm. Furthermore, it is known that DFT generally gives accurate frequencies at low computational cost[1], which enables us to treat larger systems at a high computational level.

The geometries of the molecules were optimized until the gradient was smaller than $1.0 \cdot 10^{-3}$, a frequency run was performed (without symmetry to obtain the full Hessian matrix), and the force constants obtained. Until now, the IntraFF analysis for obtaining the force field parameters was performed in a standalone program after the ADF calculation. In the near future, it will be added to the ADF program.

Results

Set of twelve molecules

The set of molecules consists of: *ethylene*, *ethane*, *methane*, *carbon monoxide*, *hydrogen*, *hydrogenchloride*, *hydrogenfluoride*, *hydrogenperoxide*, *nitrogen*, *ammonia*, *benzene*, *CFClBrH*, *dioxygen and water*. In all cases did the geometry optimization converge within five steps, which took at most one hour of CPU-time on an IBM/RS6000 per molecule. The optimized values for the *internal coordinates* of the molecules are given in Tables 3.2.1 and 3.2.2.

Probably the most commonly known experimental value for a bond angle known in computational chemistry is the one for the water molecule (104.45°). Our computed result of 104.34° is in very good agreement with this value, although it must be mentioned at the same time that the O-H bond is overestimated by some 1.6 pm. The same trend of overestimating bond lengths is seen for some other molecules, like hydrogen (2.0 pm) and ammonia (1.6 pm). On the other hand, the computed results for carbon monoxide, nitrogen and oxygen are in perfect agreement with experiment, and also the agreement for the HNH angle in ammonia is good. Generally speaking we are therefore reasonably close to the experimental geometries.

TABLE 3.2.1. FORCE CONSTANTS *FC* (KCAL/MOL/Å^2) FOR OPTIMIZED BONDS *R* (Å), WITH FROST PARAMETERS *a* (BOHR^{-1}), *b* (HARTREE) AND CORRESPONDING DISSOCIATION ENERGY D_e (EV)

Molecule	*Atoms*	*R*	*FC*	*Amber*[126]	*a*	*b*	*De*
C2H4	CC	1.324	1554	-	1.101	19.60	9.02
	CH	1.095	760	734	1.016	4.28	4.59
C2H6	CC	1.510	755	620	1.110	16.60	4.56
	CH	1.102	704	680	1.046	4.20	4.07
CH4	CH	1.098	728	680	1.035	4.24	4.28
CO	CO	1.131	2731	-	1.405	29.93	10.08
H2	HH	0.766	740	-	0.788	1.30	5.26
HCl	HCl	1.300	682	-	1.188	9.29	3.48
HF	HF	0.937	1302	-	1.435	7.09	4.29
H2O2	OO	1.443	644	-	1.548	29.02	2.21
	OH	0.981	1061	1106	1.334	6.06	4.01
N2	NN	1.098	3405	-	1.382	31.85	12.75
NH3	NH	1.024	948	868	1.178	5.20	4.42
O2	OO	1.218	1698	-	1.595	35.38	5.25
CFClBrH	CF	1.339	781	734	1.604	26.60	2.47
	CCl	1.755	457	-	1.354	37.61	2.09
	CBr	1.943	379	-	1.414	68.20	1.66
	CH	1.097	743	680	1.026	4.26	4.42
C6H6	CC	1.398	1073	938	1.133	18.34	6.43
	CH	1.090	785	734	1.029	4.20	4.20
H2O	OH	0.974	1118	1106	1.321	6.13	4.27

In order to get vibrational frequencies for the 12 molecules, we calculated the Hessian for each molecule through numerical differentiation of the analytical gradients, followed by a normal mode analysis. This procedure of numerical differentiation has the advantage that one can also obtain the derivative of the dipole moment with respect to the coordinates, and as such the IR intensities of the frequencies. The vibrational frequencies obtained from this procedure are given in Table 3.2.2. For the diatomic molecules, the values are 1582 cm^{-1} (O_2), 2167 cm^{-1} (CO), 2395 cm^{-1} (N_2), 2866 cm^{-1} (HCl), 4005 cm^{-1} (HF), 4161 cm^{-1} (H_2), with intensities 0.0 (O_2), 72.3 (CO), 0.0 (N_2), 70.3 (HCl), 145.3 (HF), 0.0 (H_2) km/mol, respectively. The frequencies are in good agreement with the experimental values[190] of 1580, 2133, 2359, 2989, 4138, 4359 cm^{-1}. For water, the DFT frequencies are 1555 [86.3], 3696 [10.6] and 3798 cm^{-1} [80.7 km/mol], with the intensities given between square brackets. The frequencies correspond more or less to one bending and two bond stretching modes. This is just a rough splitting of the *internal* coordinate systems, because in reality the normal modes do not correspond exactly to the internal coordinates, but are a mixture of them. For ammonia, we find three high frequency values (3381 [1.1], twice 3513 cm^{-1} [10.6]), and three lower values: 968 [160.5], twice 1585 cm^{-1} [0.6 km/mol]. A rough division would again lead to the high frequency modes corresponding to bond stretches, and the low values to bending modes.

TABLE 3.2.2. FORCE CONSTANTS *FC* (KCAL/MOL/RAD^2) FOR OPTIMIZED ANGLES AND DIHEDRAL ANGLES *A* (°)[a]

Angle	*A*	*FC*	*Angle*	*A*	*FC*
C2H4			*CFClBrH*		
θ_{CCH}	121.52	176.0	θ_{FCCl}	109.70	384.2
φ_{HCCH}	0/180	7.1	θ_{FCBr}	109.01	384.6
C2H6			θ_{FCH}	110.27	125.3
θ_{CCH}	111.60	155.8	θ_{ClCBr}	111.83	555.3
φ_{HCCH}	60 (± 120)	8.1	θ_{ClCH}	109.05	112.9
CH4			θ_{BrCH}	106.94	108.8
θ_{HCH}	109.47	93.8	*C6H6*		
H2O2			θ_{CCC}	120.00	735.6
θ_{OOH}	100.36	121.1	θ_{CCH}	120.00	178.7
φ_{HOOH}	106.76	6.8	*H2O*		
NH3			θ_{HOH}	104.34	89.2
θ_{HNH}	106.65	89.3			

Methane has the following frequencies: three times 1249 cm^{-1} [19.9], twice 1480 cm^{-1} [0.0], 2951 cm^{-1} [0.0] and twice 3080 cm^{-1} [7.0 km/mol]. For hydrogenperoxide, the values are 431 [150.2], 956 [0.6], 1285 [108.0], 1362 [0.9], 3618 [62.5] and 3630 cm^{-1} [24.1 km/mol]. The latter two would correspond to the OH stretches, and the first one to the HOOH torsion. The other three are some mixture of the OO stretch and the HOO bends. The bond stretches are in this case all lower than the corresponding values found in water. For ethylene, the frequencies are 798 [0.4], 924 [93.1], 937 [0.0], 1029 [0.0], 1184 [0.0], 1319 [0.0], 1390 [14.7], 1652 [0.0], 3038 [5.6], 3050 [0.0], 3117 [0.0] and 3141 cm^{-1} [5.4 km/mol]. Finally, for ethane we find 304 [0.0], twice 787 [6.6], 1019 [0.0], twice 1159 [0.0], 1332 [4.7], 1349 [0.0], twice 1425 [0.0], twice 1425 [14.3], 2949 [0.0], 2949 [42.0], twice 3010 [0.0] and twice 3033 cm^{-1} [29.2 km/mol].

The computed Hessian matrices for the molecules were used afterwards to get force constants for all possible internal coordinates of the molecules, which are given in Tables 3.2.1 and 3.2.2. Compared to the force constants of AMBER[126] (k_{OH} 1106 kcal/mol/Å^2) or GROMOS[127] ($k_{OH}^{quartic}$ 2.32·10^7 kJ/mol/nm^4) , our value for water of 1118 kcal/mol/Å^2 (or 2.46 kJ/mol/nm^4) is rather similar. Our bending constant of 89 kcal/mol/rad^2 is close to the GROMOS value (corresponding to 92 kcal/mol/rad^2), but rather far off from the AMBER value (200 kcal/mol/rad^2), which is a surprising deviation on their part. In Table 3.2.3, we report the vibrational frequencies for the molecules from the force constants obtained here, that have on average a deviation of 13 cm^{-1} from the DFT values. Here we can also see that our force constants performs well for water, which casts some more doubt on the high bending constant for water in the AMBER force field.

[a]The values of the force constants depend on how many dihedrals are taken into account in the force field; for ethane, it makes a factor 9 difference if all nine possible dihedrals are taken into account or only one; for the values reported here it is assumed that the dihedrals are all taken into account separately

TABLE 3.2.3. FREQUENCIES (CM^{-1}) FROM QUANTUMCHEMICAL AND CLASSICAL HESSIANS

QC	FF	QC	FF	QC	FF	QC	FF	QC	FF
ethylene		*ethane*		*hydrogenchloride*		*nitrogen*		*benzene*	
798	668	304	931	2866	2866	2395	2395	409[2]	462[2]
924	854	787[2]	887[2]					603[2]	885[2]
937	988	1019	1229	*hydrogenfluoride*		*oxygen*		624	744
1029	1049	1159[2]	1054[2]	4005	4005	1582	1582	716	781
1184	1186	1332	1246					805[2]	909[2]
1319	1238	1349	1450	*hydrogenperoxide*		water		931[2]	989[2]
1390	1327	1424[2]	1420[2]	431	430	1555	1537	960	896
1652	1789	1425[2]	1438[2]	956	920	3696	3703	984	972
3038	3052	2948	2921	1285	1281	3798	3757	996	1182
3050	3085	2950	2931	1362	1343			1016[2]	1054[2]
3117	3160	3010[2]	3026[2]	3618	3634	*CFClBrH*		1107	1492
3141	3166	3033[2]	3034[2]	3630	3634	215	283	1132[2]	1254[2]
						305	386	1297	1696
methane		*carbonmonoxide*		*ammonia*		419	451	1324	1730
1249[3]	1347[3]	2167	2167	968	1185	641	778	1441[2]	1644[2]
1480[2]	1508[2]			1585[2]	1593[2]	761	946	1574[2]	1909[2]
2951	2919	*hydrogen*		3381	3382	1103	1088	3119	3166
3080[3]	3085[3]	4161	4161	3513[2]	3483[2]	1154	1328	3125[2]	3169[2]
						1256	1439	3137[2]	3191[2]
						3054	3078	3148	3210

2) Twice found ;

3) Three times found

For the bonds we report in Table 3.2.1 also the parameters *a* and *b* for the Frost potential as well as the dissociation energy resulting from these parameters. These energies are in reasonable agreement with experimental values[190] (typical deviation 0.2-0.3 eV, smallest deviation 0.01 eV, largest 1.6 eV).

Conclusions

We have constructed a new method to obtain force field parameters from quantum chemical calculations for bonding interactions that can be used in molecular simulations by classical mechanics. It is invariant to the choice of internal coordinates, and can be used with any type of quantum chemical method. We have tested several options for direction vectors that are used to describe the internal coordinates, and found the one where the direction vector of maximum increase (or the gradient of the internal coordinate with respect to atomic coordinates) weighted inversely by the atomic masses to give the best performance. The heavy atoms are therefore less important for the force constant, which seems natural since they tend to move less also. The reference data come from Density Functional Theory (DFT) calculations using the LDA exchange-correlation potential in a triple zeta

valence basis set plus polarization functions (TZP), which is at a level sufficiently high to give results that are reasonably close to experimental data. From the computed DFT Hessians we have obtained the vibrational frequencies (and IR intensities) for a dozen molecules, and compared the frequencies from the force field with them. On average the difference between the two sets is 13 cm-1.

Appendix 3.2. Classical force field expressions

Harmonic potentials
The energy for any internal coordinate C using a harmonic potential is given by the following energy expression:

$$U(C_{now}) = \tfrac{1}{2} K_C \left(C_{now} - C_{eq}\right)^2 \tag{32}$$

The derivatives of this energy with respect to the atomic coordinates $r_{p,i}$ are then obtained by applying the chain rule. First, the derivative with respect to the internal coordinate is taken, then the derivative of this internal coordinate with respect to atompair distances ($R_{ij} = |r_{ij}| = |r_i - r_j|$) and finally the derivative of the atompair distance with respect to the atomic coordinates:

$$_{i,p}U(C_{now}) = G_{i,p} = \frac{\partial U(C_{now})}{\partial C}\frac{\partial C}{\partial R_{ij}}\frac{\partial R_{ij}}{\partial r_{i,p}} = K_C\left(C_{now} - C_{eq}\right)\frac{\partial C}{\partial R_{ij}}\frac{\partial R_{ij}}{\partial r_{i,p}} \tag{33}$$

The second derivatives (Hessians) are obtained by the product rule, resulting in two contributions:

$$H_{i,p,j,q} = \frac{\partial G_{i,p}}{\partial r_{j,q}} = K_C\left(\frac{\partial C}{\partial R_{ij}}\right)^2 \frac{\partial R_{ij}}{\partial r_{i,p}}\frac{\partial R_{ij}}{\partial r_{j,q}} + K_C\left(C_{now} - C_{eq}\right)\frac{\partial^2 C}{\partial R_{ij}^{\,2}}\frac{\partial R_{ij}}{\partial r_{i,p}}\frac{\partial R_{ij}}{\partial r_{j,q}} \tag{34}$$

The derivatives of a bond R with respect to the atomic coordinates are given by:

$$\begin{aligned}
\frac{\partial R_{ij}}{\partial r_{i,p}} &= \frac{r_{ij,p}}{R_{ij}}\frac{\partial r_{ij}}{\partial r_i} \\
\frac{\partial^2 R_{ij}}{\partial r_{i,p}\partial r_{j,q}} &= (-1)^{1+\delta_{ij}}\left[\frac{\delta_{pq}}{R_{ij}} - \frac{r_{ij,p}\,r_{ij,q}}{R_{ij}^{\,3}}\right]
\end{aligned} \tag{35}$$

For the dihedral angle ϕ, we use the formulations of Blondel and Karplus[191], who have shown that for any normal dihedral angle value (i.e. all cases except the one where three or more atoms lie on a straight line), the derivatives for the potential of ϕ_{ijkl} can be obtained by using the following vectors:

$$\begin{aligned}
V_F &= r_i - r_j \\
V_G &= r_j - r_k \\
V_H &= r_l - r_k
\end{aligned}
\qquad
\begin{aligned}
V_A &= V_F \times V_G \\
V_B &= V_H \times V_G
\end{aligned} \tag{36}$$

The derivatives for the four atoms are then given by:

$$\begin{aligned}
\frac{\partial\phi}{\partial r_i} &= -\frac{|V_G|}{\left(V_A{}^T V_A\right)} V_A \\
\frac{\partial\phi}{\partial r_j} &= +\frac{|V_G|}{\left(V_A{}^T V_A\right)} V_A + \frac{\left(V_F{}^T V_G\right)}{\left(V_A{}^T V_A\right)|V_G|} V_A - \frac{\left(V_H{}^T V_G\right)}{\left(V_B{}^T V_B\right)|V_G|} V_B \\
\frac{\partial\phi}{\partial r_k} &= -\frac{|V_G|}{\left(V_B{}^T V_B\right)} V_B - \frac{\left(V_F{}^T V_G\right)}{\left(V_A{}^T V_A\right)|V_G|} V_A + \frac{\left(V_H{}^T V_G\right)}{\left(V_B{}^T V_B\right)|V_G|} V_B \\
\frac{\partial\phi}{\partial r_l} &= +\frac{|V_G|}{\left(V_B{}^T V_B\right)} V_B
\end{aligned} \tag{37}$$

From these equations, one can easily see why this procedure breaks down if three atoms lie on a straight line, since the inner product of either the V_A- or V_B-vector will be zero. However, in general such a situation is not likely to occur, since there are bending interactions preventing it from happening. The second derivatives of the dihedral angle with respect to the atomic coordinates are given by first taking the derivatives with respect to the V_F, V_G, V_H -vectors:

$$\begin{aligned}
\frac{\partial^2\phi}{\partial V_F{}^2} &= \frac{|V_G|}{\left(V_A{}^T V_A\right)^2}\left(V_A \quad V_G \times V_A + V_G \times V_A \quad V_A\right) \\
\frac{\partial^2\phi}{\partial V_H{}^2} &= \frac{-|V_G|}{\left(V_B{}^T V_B\right)^2}\left(V_B \quad V_G \times V_B + V_G \times V_B \quad V_B\right) \\
\frac{\partial^2\phi}{\partial V_F \partial V_G} &= \frac{1}{|V_G|\left(V_A{}^T V_A\right)^2}\left(\left(V_G{}^T V_G\right)\left(V_A \times V_F\right) \quad V_A + \left(V_F{}^T V_G\right) V_A \quad \left(V_A \times V_G\right)\right) \\
\frac{\partial^2\phi}{\partial V_G \partial V_H} &= \frac{-1}{|V_G|\left(V_B{}^T V_B\right)^2}\left(\left(V_G{}^T V_G\right)\left(V_B \times V_H\right) \quad V_B + \left(V_H{}^T V_G\right) V_B \quad \left(V_B \times V_G\right)\right) \\
\frac{\partial^2\phi}{\partial V_F \partial V_H} &= 0
\end{aligned} \tag{38}$$

$$\begin{aligned}
\frac{\partial^2\phi}{\partial V_G{}^2} = {} & \frac{1}{2|V_G|^3\left(V_A{}^T V_A\right)}\left(V_G \times V_A \quad V_A + V_A \quad V_G \times V_A\right) \\
& + \frac{\left(V_F{}^T V_G\right)}{|V_G|\left(V_A{}^T V_A\right)^2}\left(V_A \quad V_F \times V_A + V_F \times V_A \quad V_A\right) \\
& - \frac{1}{2|V_G|^3\left(V_B{}^T V_B\right)}\left(V_G \times V_B \quad V_B + V_B \quad V_G \times V_B\right) \\
& - \frac{\left(V_H{}^T V_G\right)}{|V_G|\left(V_B{}^T V_B\right)^2}\left(V_B \quad V_H \times V_B + V_H \times V_B \quad V_B\right)
\end{aligned} \tag{39}$$

where $\quad$ stands for the tensorproduct:

$$\left(A \quad B\right)_{ij} = A_i B_j \tag{40}$$

Now, with the derivatives of the V_F, V_G, V_H -vectors with respect to the atomic coordinates:

$$\frac{\partial V_F}{\partial r_i} = 1\ ;\ \frac{\partial V_F}{\partial r_j} = -1\ ;\ \frac{\partial V_G}{\partial r_j} = 1\ ;\ \frac{\partial V_G}{\partial r_k} = -1\ ;\ \frac{\partial V_H}{\partial r_l} = 1\ ;\ \frac{\partial V_H}{\partial r_k} = -1 \tag{41}$$

we finally obtain the second derivatives of the dihedral angle as:

$$\frac{\partial^2 \phi}{\partial r_i^{\,2}} = \sum_{p,q} \frac{\partial^2 \phi}{\partial V_p \partial V_q} \frac{\partial V_p}{\partial r_i} \frac{\partial V_q}{\partial r_i} \quad ; p,q \quad \{F,G,H\} \tag{42}$$

By following the same procedure for the angle, and by using the following vectors:

$$\begin{aligned} V_A &= r_i - r_j \\ V_B &= r_k - r_j \end{aligned} \qquad V_G = V_A \times V_B \tag{43}$$

we obtain the derivative of the angle θ with respect to the atomic coordinates as follows:

$$\begin{aligned} \frac{\partial \theta}{\partial V_A} &= \frac{-1}{\left(V_A{}^T V_A\right)\left|V_G\right|}\left(V_G \times V_A\right) \\ \frac{\partial \theta}{\partial V_B} &= \frac{-1}{\left(V_B{}^T V_B\right)\left|V_G\right|}\left(V_B \times V_G\right) \end{aligned} \qquad \begin{aligned} \frac{\partial \theta}{\partial r_i} &= \frac{\partial \theta}{\partial V_A} \\ \frac{\partial \theta}{\partial r_j} &= -\left(\frac{\partial \theta}{\partial V_A} + \frac{\partial \theta}{\partial V_B}\right) \\ \frac{\partial \theta}{\partial r_k} &= \frac{\partial \theta}{\partial V_B} \end{aligned} \tag{44}$$

The second derivative has not been worked out yet, when needed it is obtained by numerical differentiation of eq. (44).

Anharmonic Frost potential[131,132]

The energy expression for this anharmonic bond potential is for a bond R_{ij} between atoms i and j as:

$$U\left(R_{ij}\right) = e^{-aR_{ij}}\left(\frac{N_i N_j}{R_{ij}} - b\right) \tag{45}$$

The derivatives with respect to the distance R_{ij} are given by:

$$\begin{aligned} G\left(R_{ij}\right) &= e^{-aR_{ij}}\left(ab - \frac{aN_iN_j}{R_{ij}} - \frac{N_iN_j}{R_{ij}^{\,2}}\right) \\ H\left(R_{ij}\right) &= e^{-aR_{ij}}\left(\frac{2N_iN_j}{R_{ij}^{\,3}} + \frac{2aN_iN_j}{R_{ij}^{\,2}} + \frac{a^2N_iN_j}{R_{ij}} - a^2b\right) \end{aligned} \tag{46}$$

Quartic GROMOS potential[127]

In order to avoid the square root operation, GROMOS96 uses quartic potentials for bonds. The expressions for a bond R are:

$$\begin{aligned} U(R) &= \tfrac{1}{4} K^{quartic}\left(R^2 - R_{eq}{}^2\right)^2 \\ G(R) &= K^{quartic} R\left(R^2 - R_{eq}{}^2\right) \\ H(R) &= K^{quartic}\left(3R^2 - R_{eq}{}^2\right) \end{aligned} \tag{47}$$

Dihedral potentials

For the dihedral angle, one normally takes either a harmonic potential, or a periodic one (with periodicity n):

$$\begin{aligned} U(\phi_{now}) &= K_\phi\left(1 + \cos\left(n\phi_{now} - \phi_{shift}\right)\right) = K_\phi\left(1 + \cos\left(n\left(\phi_{now} - \phi_{eq}\right) - \pi\right)\right) \\ \phi_{shift} &= n\phi_{eq} + \pi \end{aligned} \tag{48}$$

The derivatives with respect to the dihedral angle are then:

$$\begin{aligned} G(\phi_{now}) &= -nK_\phi \sin\left(n\phi_{now} - \phi_{shift}\right) \\ H(\phi_{now}) &= -n^2 K_\phi \cos\left(n\phi_{now} - \phi_{shift}\right) \end{aligned} \tag{49}$$

chapter four

APPLICATIONS *of Density Functional Theory*

4

CONTENTS

SUMMARY

The application of Density Functional Theory to three different chemical topics is discussed in Chapter 4. First the computation of molecular polarizabilities and the influence of the choice for the basis set and exchange-correlation potential on the accuracy is discussed; then the accuracy of optimized geometries of several exchange-correlation potentials in a number of basis sets is presented, where the test set consists either of a set of small molecules that was used previously by others to check the accuracy of several wavefunction based methods. Finally, the new charge analysis presented in Chapter 3 is validated by checking its use for the concepts of molecular recognition, electron withdrawing-donating groups and electrophilic substitution reactions.

4.1 *Polarizabilities*

A comparison of Restricted Hartree Fock, Density Functional Theory and Direct Reaction Field mean values for organic molecules

There is a growing interest in atomic and molecular polarizabilities within the scope of the development of accurate force fields to be used in QM/MM methods[133,134,138,192-194]. Dispersion and induction forces are more and more being recognized as essential parts in the description of intermolecular interactions in chemical environments[138,195-201].

It has also been noticed that using a molecular polarizability located in the center of the molecule, often leads to improper behavior[138,202]. This can be overcome by using effective atomic polarizabilities that represent the molecular value. A set of these atomic parameters have been constructed within the framework of the Direct Reaction Field (DRF) method, which predict molecular polarizabilities with a deviation comparable to the estimated experimental uncertainty[203,204] (2-4 %). In this study, we present DRF results for a representative series of organic molecules. For comparison, we have also performed Restricted Hartree Fock (RHF) and Time Dependent Density Functional Theory (TD-DFT) calculations on the same set of molecules, the latter also at non-zero frequency.

The computation of a polarizability α is based on a Taylor expansion of the total energy about the electric field strength E:

$$U = U^{(0)} - \mu_i^{(0)} E_i - \frac{1}{2!}\alpha_{ij} E_i E_j - \frac{1}{3!}\beta_{ijk} E_i E_j E_k - \frac{1}{4!}\gamma_{ijkl} E_i E_j E_k E_l - \ldots \quad (1)$$

with $U^{(0)}$ the unperturbed total energy, $\mu^{(0)}$ the permanent dipole moment, α the polarizability and β and γ the first and second hyperpolarizabilities. The appropriate derivatives of the energy with respect to the electric field then result in expressions for the (hyper)polarizabilities. This derivative can be obtained either analytically or numerically by a finite field procedure. The former is faster but not standard available for all kinds of quantumchemical methods, while the latter requires only the calculation of the energy or the dipole moment, but is in general much slower, since the energy/dipole needs to be calculated many times.

Methods

Restricted Hartree Fock

The RHF results are obtained with Coupled Perturbative Hartree Fock (CPHF) equations[204] (as implemented in the HONDRF program[133]), which take the response of a molecule to an applied field analytically into account. Dunning's triple zeta valence plus polarization functions basis set (TZP)[203] was used for all molecules, which is the largest standard (Gaussian) basis set in the HONDRF program and has proven to give an underestimation of the polarizabilities by ~ 22 %[139]. A lot of effort has been put into constructing basis sets especially designed for accurate calculations of the polarizability[205-209]. However, they are based on Gaussian type orbitals, which makes a comparison with the ADF results (using Slater type orbitals) impossible. Moreover, it is not the purpose of this study to compare *basis*

sets or to construct them for the ADF program, but rather to make an honest comparison between the three *methods* using the programs and basis sets that are readily available.

Density Functional Theory
The DFT results were all obtained by using the RESPONSE code[210-212] in the Amsterdam Density Functional (ADF) program[117,177,213-215]. This Time Dependent DFT (TD-DFT) method is at zero frequency similar to a CPHF procedure, but then applied to DFT SCF equations which gives the same polarizabilities as when using a finite field procedure. The ADF program uses basis sets of Slater functions, of which two were used: a triple zeta valence plus polarization (TZP, basis set IV in ADF; of equal size as Dunning's TZP basis) and a special one with triple zeta valence plus double polarization functions and diffuse s, p, d functions (TZ2P++). This basis set has been constructed for improving the results and has been reported to give accurate polarizabilities[210]. Several potentials were used: LDA (VWN[119]) and GGA's (Becke88 exchange[120]-Perdew86 correlation[121] potential, Van Leeuwen-Baerends LB94[216]).

Direct Reaction Field
Within the DRF approach[133], a molecular polarizability is being constructed from interacting atomic polarizabilities α_p:

$$\mu_p = \alpha_p \left(E_0 + \sum_{q \neq p}^{N} T_{pq} \mu_q \right) \tag{2}$$

with the *modified* dipole field tensor T_{pq}:

$$\left(T_{pq}\right)_{ij} = \left(3\mathbf{r}_i \mathbf{r}_j f_T - \delta_{ij} r^2 f_E \right) / r^5 \tag{3}$$

This can be written also as a matrix equation:

$$\mathbf{M} = \alpha(\mathbf{E} + \mathbf{TM}) \tag{4}$$

which can be solved directly with the matrix $\mathbf{A} = \alpha^{-1} - \mathbf{T}$, by using the inverse of matrix $\mathbf{A}$:

$$\mathbf{AM} = \mathbf{E} \rightarrow \mathbf{M} = \mathbf{A}^{-1}\mathbf{E} \tag{5}$$

This so-called relay matrix $\mathbf{R} = \mathbf{A}^{-1}$ gives the linear response of the molecule to a given external field, i.e. its polarizability in a $3N x 3N$ representation. It can be reduced to a "normal" $3x3$-tensor, resulting in the molecular polarizability tensor:

$$\alpha_{mn} = \sum_{i,j=1}^{N} \left(\mathbf{R}_{ij} \right)_{mn} \quad ; \; m,n \quad \{x,y,z\} \tag{6}$$

The modification is present in the screening factors f_T and f_E, which represent the damping due to overlapping charge densities. Several functions have been tested for this damping[217], of which one has survived:

$$\rho(u) = \frac{a^3}{8\pi} e^{-au} \qquad u = r_{ij} \Big/ \left(\alpha_i \alpha_j\right)^{1/6} \qquad v = au$$

$$f_V = 1 - \left(\tfrac{1}{2} v + 1\right) e^{-v}$$
$$f_E = f_V - \left(\tfrac{1}{2} v^2 + \tfrac{1}{2} v\right) e^{-v} \qquad (7)$$
$$f_T = f_E - \tfrac{1}{6} v^3 e^{-v}$$

In these equations the a-factor and the atomic polarizabilities are adjustable parameters obtained in the fit procedure[139] (see Table 4.1.1).

TABLE 4.1.1. DRF ATOMIC POLARIZABILITIES α (BOHR^3) WITH A-FACTOR 2.1304

atom	α	*atom*	α	*atom*	α
H	2.793	O	5.749	Cl	16.198
C	8.696	S	16.698	Br	23.571
N	6.557	F	3.001	I	36.988

Computational details

The DRF and RHF calculations were carried out on the Cray J932 supercomputer in Groningen. All ADF results were calculated at a SGI PowerChallenge workstation. The geometries of the molecules were taken as much as possible from experimental data[218]. The remaining (internal) coordinates were optimized by the PM3 method in MOPAC93[219].

Results and discussion

The calculated mean polarizabilities of the 15 molecules are given in Table 4.1.2, together with experimental data. The experimental mean values are obtained mostly from the refractive index n (at 5893 Å ; sodium D-line) and the Lorentz-Lorentz equation (with M molecular weight, ρ macroscopic density, N_{av} Avogadro's number):

$$\frac{n^2 - 1}{n^2 + 2} \frac{M}{\rho} = \frac{4\pi}{3} N_{av} \bar{\alpha} \qquad (8)$$

These values should be extrapolated to infinite wavelength (or zero frequency) to obtain the static polarizability. This scales the values down by 1-4 %[220], and gives an estimate of the experimental uncertainty when using the uncorrected values.

The experimental data can also be obtained from the Kerr constants[220], but this introduces, apart from the wavelength dependency, an additional uncertainty. One has to make an assumption of the geometry, thereby limiting this method to (small) symmetric molecules. The additional uncertainty has been estimated to be 5-10 %[220], giving a total uncertainty of at least 6-14 %.

In Table 4.1.3, some experimental polarizability components are given, which were taken from references 220 and 221. Some other components were left out, because of spurious

assumptions that increase the uncertainty even more. If necessary, the components were uniformly scaled to reproduce the average mean polarizability. It should be noted also, that some *experimental* components reported by Applequist et al.[221] are extrapolated empirical values; i.e. experimental Kerr constants were obtained for some molecules which were then extrapolated to similar (larger) molecules.

TABLE 4.1.2. MOLECULAR POLARIZABILITIES (BOHR^3) USING TZP BASIS SET

	$\alpha_{exp}\nu$		α_{DRF}	α_{RHF}	α_{LDA}	α_{BP}	α_{LB94}
Acetamide^p	40.5	Rl	38.6	31.1	40.5	39.5	39.2
Acetylene^t	22.5	Kg	21.9	18.0	19.6	19.3	19.2
Benzene^p	70.1	Rl	61.9	61.9	66.7	65.7	65.9
Chlorine^p	31.1	Kg	31.2	18.7	25.3	24.9	25.4
Cyclohexanol^t	79.9	Rl	78.0	69.2	79.2	77.0	77.0
Dimethylether^p	35.0	Rg	35.4	27.6	34.8	33.8	33.4
Formaldehyde^p	16.5	Rg	18.3	13.5	16.6	16.3	16.1
Hydrogen^p	5.3	Kg	4.9	2.6	4.8	4.6	4.7
Methylcyanide^p	29.7	Rl	29.8	24.6	28.4	27.8	28.2
Neopentane^t	69.0	Rl	65.5	59.0	67.8	65.7	65.7
Propane^p	42.4	Rl	42.2	35.9	42.1	40.8	41.5
TCFM^t	57.5	Rl	56.5	40.0	53.2	52.8	52.7
TCMC^t	70.5	Rl	68.3	54.0	66.1	65.0	66.1
TFM^p	19.0	Kg	19.0	13.1	17.9	17.5	16.8
Water^p	9.94	Rl	10.1	5.6	8.4	8.3	8.0
Time			~ 1 s	57 hr	9 hr	14 hr	23 hr
Average deviation (%)			-1.8	-24.9	-6.0	-8.1	-8.2
			± 4. 8	*± 11.5*	*± 5.8*	*± 5.5*	*± 5.5*
Average absolute dev. (%)			3.6	24.9	6.1	8.1	8.2
			± 3.6	*± 11.5*	*± 5.7*	*± 5.5*	*± 5.5*
Relative to LB94 (%)					+2.4	+0.1	
					± 1.9	*± 2.0*	

TCFM	trichlorofluoromethane
TCMC	trichloromethylcyanide
TFM	trifluoromethane
p	Molecules used to obtain DRF-parameters[139]
t	Molecules used to test DRF-parameters
K	Obtained from Kerr constants (uncertainty ca. 5 %)
R	Obtained from refractive index (uncertainty ca. 0.2 %)
g	Gas phase
l	Liquid phase
DRF	Direct Reaction Field approach result[139]
RHF	Restricted Hartree Fock
LDA	Local Density Approximation result (Vosko-Wilk-Nusair potential[119])
BP	Becke88 exchange[120] with Perdew correlation[121] potential
LB94	Van Leeuwen-Baerends (LB94) potential with correct asymptotic behavior[216]
ν	Polarizabilities at 5893 Å

Restricted Hartree Fock

The RHF results give an average deviation of 24.9 %, which has been reported before[139], and can be attributed to two factors. The first is the (small) basis set ; it is known that specially constructed[205-209] or large (including very diffuse functions)[210] basis sets are needed to obtain accurate results. This is especially apparent from the results of planar and linear molecules, where the out-of-plane polarizability is very small in comparison to the experimental and DFT results.

The second factor is the absence of electron correlation in the RHF method. Inclusion of this correlation (by for instance Configuration Interaction, Multi Configuration SCF, Coupled Cluster or Møller-Plesset methods) increases the work needed enormously, and can be applied only to relatively small molecules.

TABLE 4.1.3. POLARIZABILITY COMPONENTS[#] (BOHR[3])

	α_{exp}			α_{DRF}			α_{RHF}			α_{LB94}		
	xx	yy	zz	xx	yy	zz	xx	yy	zz	xx	yy	zz
Acetylene	18.5	18.5	30.5	15.1	15.1	35.6	11.5	11.5	31.0	17.7	17.7	31.8
Benzene	82.7	82.7	45.0	74.9	74.9	35.8	74.8	74.8	36.0	83.7	83.7	41.7
Chlorine	24.4	24.4	44.5	27.3	27.3	39.0	11.0	11.0	34.2	24.8	24.8	40.0
Dimethylether	33.0	29.3	42.6	34.1	32.6	39.4	25.9	25.7	31.2	32.9	32.8	41.0
Hydrogen	4.9	4.9	6.3	4.1	4.1	6.5	0.7	0.7	6.5	4.9	4.9	7.0
Methylcyanide	25.5	25.5	38.1	24.9	24.9	39.7	19.2	19.2	35.4	25.0	25.0	42.4
Neopentane	69.0	69.0	69.0	65.5	65.5	65.5	59.0	59.0	59.0	68.6	68.6	68.6
TCFM	59.7	59.7	53.0	60.0	60.0	49.4	44.3	44.3	31.4	63.3	63.3	51.3
TCMC	69.6	69.6	72.3	67.1	67.1	70.6	53.0	53.0	56.0	68.2	68.2	84.6
TFM	19.4	19.4	18.1	18.8	18.8	19.2	13.4	13.4	12.5	19.1	19.1	17.6
Av. dev. (%)				-3.4 ± *9.4*			-27.7 ± *20.5*			0.9 ± *7.2*		
Av.abs. dev. (%)				8.0 ± *5.9*			28.1 ± *20.1*			5.0 ± *5.2*		

#	the molecules are oriented with the main symmetrical axis along the z-axis
TCFM	trichlorofluoromethane
TCMC	trichloromethylcyanide
TFM	trifluoromethane
exp	experimental polarizability components[220,221] (accuracy 6-14 %)
DRF	Direct Reaction Field approach result[139]
TZP	CPHF value with TZP basis set
VII	Van Leeuwen-Baerends (LB94) potential[216] in basis set VII (TZ2P++)

Density Functional Theory

Much more promising and accurate are the DFT results. The results from all three potentials in the TZP basis set give a much better accuracy (dev. 6-8 %; Table 4.1.2) for the mean values. When using the TZ2P++ basis set (Table 4.1.4), which has been reported as the basis set limit for (other) organic molecules[210], the deviations become even smaller (LDA 5.8 %, Becke-Perdew 3.6 %, LB94 potential 3.0 %). In both basis sets, the LB94 results are smaller than either the LDA or the Becke-Perdew results. The DFT polarizability components show, like the mean value, a much smaller deviation from the experimental data than the RHF results, and is far less than the experimental uncertainty of these values.

However, we are comparing with the *uncorrected* mean values. A better comparison can then be made by calculating the (frequency dependent) polarizabilities with the frequency corresponding to the wavelength (5893 Å) of the sodium D-line. The frequency dependent polarizabilities from the LB94 potential in the TZ2P++ basis set are on average 2.9 % larger than the static values, which is in good agreement with the estimated 1-4 % from experimental data[220]. They are also on average 4.7 % larger than the experimental values, which is a deviation comparable to the experimental uncertainty.

TABLE 4.1.4. MOLECULAR POLARIZABILITIES (BOHR^3) USING TZ2P++ BASIS SET

	α_{LDA}	α_{BP}	α_{LB94}	α_{LB94}^{ν}
Acetamide	43.7	42.6	41.2	43.1
Acetylene	23.8	23.5	22.4	23.1
Benzene	70.8	69.6	69.7	72.6
Chlorine	32.0	31.3	29.9	31.7
Cyclohexanol	82.1	79.9	80.7	82.7
Dimethylether	37.2	36.0	35.6	36.3
Formaldehyde	19.1	18.6	17.8	18.3
Hydrogen	5.9	5.5	5.6	5.8
Methylcyanide	31.2	30.5	30.8	31.1
Neopentane	69.5	67.4	68.6	70.3
Propane	43.7	42.3	42.1	44.1
TCFM	60.9	60.0	59.3	60.9
TCMC	74.6	73.3	73.7	75.2
TFM	20.7	20.3	18.6	18.9
Water	10.6	10.3	9.2	9.4
Time	32 hr	53 hr	80 hr	80 hr
Av. dev. (%)	5.84	3.17	0.89	3.79
	± 3.70	*± 3.55*	*± 3.73*	*± 3.69*
Av. abs. dev. (%)	5.84	3.62	2.98	4.65
	± 3.70	*± 3.09*	*± 2.41*	*± 2.52*
Rel. to LB94 (%)	+4.98	+2.35		+2.90
	± 3.75	*± 3.91*		*± 1.35*

TCFM	trichlorofluoromethane
TCMC	trichloromethylcyanide
TFM	trifluoromethane
LDA	Local Density Approximation result (Vosko-Wilk-Nusair potential[119])
BP	Becke88 exchange[120] with Perdew correlation[121] potential
LB94	Van Leeuwen-Baerends (LB94) potential with correct asymptotic behavior[216]
ν	polarizabilities at 5893 Å

In principle, both CPHF and TD-DFT should scale as N^3, where N is the number of basis functions. While the CPHF method (taking about 25 % of the total CPU time) indeed practically scales as N^3, the TD-DFT method more or less scales linearly and takes approximately 40 % of the CPU time. Apparently the current systems are too small for the N^3 behavior of TD-DFT to become dominant.

Direct Reaction Field

The DRF approach gives polarizabilities with an accuracy of 3.6 %, at a very low computational cost (< 1s) and with high transferability to other molecules[139]. It should be noticed that, like the RHF method, some problems arise with linear and planar molecules, where the anisotropy of the molecular polarizability (i.e. π-bonds) cannot be obtained with interaction tensors based on the atomic positions. Adding extra fit points increases the accuracy of the polarizabilities[202], but leads also to bad interaction energies in classical DRF energy calculations[138].

Conclusions

The *DRF* approach provides mean polarizability values at low computational cost with an accuracy (3-4 %) equal to experimental uncertainty[139]. Its only setback is the underestimation of the anisotropy of linear and planar molecules, but that is more than compensated with a high transferability to other molecules without the need to reparameterize.

The *RHF* method gives rather poor mean polarizabilities (deviation 25 % in TZP basis), with high cost for improvement upon these results. Even with the specially constructed (polarized) basis sets, one needs correlated wavefunctions, thereby limiting the applicability to small molecules.

The *DFT* methods give good mean values with the TZP basis set (6-8 % deviation), and accurate values (dev. 3-6 %) with the TZ2P++ basis set. The latter needs only 3-4 times more CPU time, so it is rather easy to improve the TZP results. It is also evident that the LDA gives larger values than the gradient corrected potentials (+2-5 %), while the Becke-Perdew results are substantially (2.4 %) improved by the Van Leeuwen-Baerends potential.

The impact of the frequency dependency on the polarizabilities is reflected properly by the TD-DFT method which was shown with the Van Leeuwen-Baerends potential. The frequency dependent polarizabilities are on average 2.9 % larger than the static values which is in perfect agreement with the extrapolation estimate of 1-4 %.

The polarizability anisotropy is very well reproduced by all *DFT* methods, and the *DRF* method, giving deviations (5-12 %) from experimental values within the experimental uncertainty (6-14 %). The *RHF* method again shows deviations (28 %) which are large in comparison to experimental deviations.

4.2 *Accuracy of geometries*

The influence of basis sets, xc potentials, core electrons and relativistic corrections on geometries of small molecules and metallocenes

Before one can perform a quantum chemical calculation, one needs to have a structure of the molecule; e.g. one needs to know where the atoms are positioned in space (its geometry). The accuracy with which these geometries can be predicted by different quantum chemical methods is a very useful thing to know, allowing an estimate of the reliability of a computed geometry.

In a recent paper, Helgaker et al.[188] presented a systematic investigation of the accuracy obtainable with wavefunction based methods, applying a hierarchy of basis sets and methods on a set of 19 small closed-shell molecules[a]. They used Hartree-Fock (HF), Configuration Interaction (CISD), Møller-Plesset (MP2, MP3, MP4) and coupled cluster methods (CCSD, CCSD(T)) in Dunning's correlation consistent basis sets (cc-pVDZ, cc-pVTZ, cc-pVQZ)[222], and looked at the mean error, standard deviation, mean absolute error and maximum error. Hartree Fock is shown to result always in a too short bond distance, while the inclusion of the electron correlation tends to increase it. Increasing the basis set size tends to decrease the bond distance again. The best results are obtained when one uses the CCSD(T) method, either in the cc-pVTZ or the cc-pVQZ basis, with an accuracy (0.0022 Å) that is comparable to the experimental uncertainty. In another paper, Helgaker et al.[223] investigated the molecular structure of ferrocene. That study showed clearly the limitations of using the CCSD or CCSD(T) methods. They could not calculate the gradients with these methods, and therefore had to *estimate* the equilibrium Fe-ring distance in ferrocene based on three single-point energy calculations.

In this study, a similar systematic investigation is presented where the accuracy of geometries as predicted by Density Functional Theory is studied. The influence of the basis set, treatment of core electrons and relativistic corrections has been investigated for several currently available exchange-correlation potentials. The same set of 19 small molecules is used in the first part, to enable not only a comparison with experimental data but also with the wavefunction based results obtained by Helgaker et al. For these molecules, the influence of including the core electrons in the optimizations has been investigated. Besides this set of molecules, also a set of metallocene molecules has been used[b], where not only the inclusion of core electrons was investigated, but also the effect of scalar relativistic corrections (with the ZORA[117,224] hamiltonian). All of these calculations were performed in the standard available basis sets, ranging from a single zeta valence basis set (SZV, I) to a triple zeta valence basis set plus double polarization functions (TZ2P, VII). As B3LYP[122] is not a *pure* DFT potential, and needs a portion of Hartree-Fock exchange, it can not be used in the ADF program. To check the accuracy of it, as well as to compare the ADF basis sets to the basis sets used in the paper by Helgaker et al.[188], the geometry optimizations were performed also with the B3LYP and BLYP potentials in the cc-pVDZ and cc-pVTZ basis sets using the HONDO98[204,225] program.

---

[a] The set of small molecules consists of: HF, H_2O, NH_3, CH_4, N_2, CH_2, CO, HCN, CO_2, HNC, C_2H_2, CH_2O, HNO, N_2H_2, O_3, C_2H_4, F_2, HOF, H_2O_2.

[b] The metallocene set: manganocene, ferrocene, cobaltocene, nickelocene and ruthenocene.

Computational details

In the ADF program[117,187], standard basis sets are available ranging from small (SZV, I) to large (TZ2P, VII), with for the basis sets up to TZ2P (V) the option of either to include (all electron) or exclude (frozen core) the core electrons explicitly in the calculations. In the frozen core basis sets, there are still basis functions assigned to the core electrons; the basis functions of the valence electrons are then explicitly orthogonalized to them. As the calculations are significantly faster when the core electrons are not included, this is normally the preferred option. The following standard available exchange-correlation potentials were examined: Local Density Approximation (LDA), Becke88 exchange[120] combined with Perdew86 correlation[121] (Becke-Perdew), BLAP3[226], Becke88 exchange[120] with Lee-Young-Parr correlation[227] (BLYP), Perdew-Burke-Ernzerhof[228] (PBE), Perdew-Wang (PW91)[189,229], Revised Perdew-Burke-Ernzerhof[230] (REVPBE) and RPBE[231].

To compare the basis sets, the cc-pVDZ and cc-pVTZ basis sets were used also for the BLYP potential using the HONDO98 program. Moreover, also the B3LYP potential has been used with this program in the same basis sets to enable a rough comparison with the *pure* DFT xc-potentials.

Scalar relativistic corrections can be included in the calculations quite easily in ADF, using the Zeroth Order Regular Approximation (ZORA)[117,224], which is generally found to give an accurate description of the relativistic effects. Although spin-orbit coupling is possible also with the ZORA approach, these effects were not included since the gradients are not implemented yet; therefore it is not possible to optimize the geometries directly.

TABLE 4.2.1. EXPERIMENTAL BOND LENGTHS (pm)

Molecule	Bond	Bond length	Molecule	Bond	Bond length
C_2H_2	CH	106.2	HCN	CH	106.5
C_2H_2	CC	120.3	HCN	CN	115.3
C_2H_4	CH	108.1	HF	HF	91.7
C_2H_4	CC	133.4	HNC	NH	99.4
CH_2	CH	110.7	HNC	CN	116.9
CH_2O	CH	109.9	HNO	NH	106.3
CH_2O	CO	120.3	HNO	NO	121.2
CH_4	CH	108.6	HOF	OH	96.6
CO	CO	112.8	HOF	OF	143.5
CO_2	CO	116.0	N_2	NN	109.8
F_2	FF	141.2	N_2H_2	NH	102.8
H_2O	OH	95.7	N_2H_2	NN	125.2
H_2O_2	OH	96.7	NH_3	NH	101.2
H_2O_2	OO	145.6	O_3	OO	127.2

Several statistical measures have been used to quantify the accuracy of the methods. The difference between the calculated (R_i^{calc}) and experimental (R_i^{exp}) bond length gives the error Δ_i:

$$\Delta_i = R_i^{calc} - R_i^{exp} \quad (1)$$

For each basis set (if possible, both with and without including the core electrons explicitly) and exchange-correlation potential, the mean error $\bar{\Delta}$, the standard deviation in the errors Δ_{std}, the mean absolute error $\bar{\Delta}_{abs}$ and the maximum error Δ_{max} were calculated:

$$\bar{\Delta} = \frac{1}{n}\sum_{i=1}^{n}\Delta_i \qquad \Delta_{std} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\Delta_i - \bar{\Delta}\right)^2} \qquad \bar{\Delta}_{abs} = \frac{1}{n}\sum_{i=1}^{n}\left|\Delta_i\right| \qquad \Delta_{\max} = \max\left|\Delta_i\right| \tag{2}$$

Each measure characterizes a specific aspect of the performance of the xc-potentials and basis sets. The two first measures characterize the distribution of errors about a mean value $\bar{\Pi}$ for a given xc-potential in a certain basis set, thus quantifying both systematic and non-systematic errors. The mean absolute error represents the typical magnitude of the errors in the calculations, while the maximum error indicates how large the errors can be.

Small molecules

The experimental values for the bond lengths of the set of small molecules are given in Table 4.2.1, while the mean errors for the eight exchange-correlation potentials in basis sets I to VII are given in Table 4.2.2, both in an all-electron and frozen core basis if available.

TABLE 4.2.2. MEAN ERRORS (pm)

XC-potential	I	II	III	IV	V	VI	VII
frozen core							
Becke-Perdew	5.58	4.00	1.49	1.24	0.90	-	-
BLAP3	6.27	4.16	1.63	1.38	1.02	-	-
BLYP	6.09	4.32	1.82	1.55	1.21	-	-
LDA	4.65	3.26	0.83	0.61	0.28	-	-
PBE	5.44	3.97	1.43	1.19	0.87	-	-
PW91	5.47	3.81	1.31	1.08	0.74	-	-
REVPBE	5.75	4.28	1.68	1.45	1.13	-	-
RPBE	5.76	4.36	1.76	1.54	1.21	-	-
average	*5.63*	*4.02*	*1.49*	*1.26*	*0.92*	-	-
all electron							
Becke-Perdew	5.54	3.96	1.44	1.30	0.98	0.81	0.82
BLAP3	6.20	4.09	1.58	1.44	1.09	0.92	0.91
BLYP	6.00	4.24	1.76	1.61	1.26	1.09	1.08
LDA	4.55	3.16	0.74	0.62	0.31	0.17	0.17
PBE	5.36	3.93	1.37	1.29	0.93	0.78	0.78
PW91	5.38	3.76	1.25	1.14	0.82	0.66	0.65
REVPBE	5.65	4.23	1.66	1.52	1.20	1.04	1.04
RPBE	5.76	4.33	1.74	1.61	1.28	1.12	1.12
average	*5.56*	*3.96*	*1.44*	*1.32*	*0.98*	*0.82*	*0.82*

In all cases, the bond lengths are on average overestimated, with an improvement of the results as the basis set size is increased. For the frozen core basis sets, the average of the

mean errors for the eight xc-potentials improves gradually from 5.63 pm in basis I to 0.92 pm in basis set V. The same trend is observed for the all electron basis sets, which show still an improvement on going from basis set V to VI of some 0.2 pm. However, going from basis set VI to VII, no further improvement is observed, which indicates that the basis set limit has been reached (at least concerning the accuracy of geometries). The difference between the all electron and frozen core basis sets are quite small; the all electron basis sets seem to perform slightly better in the smaller basis sets, while the frozen core seems to give slightly better results in basis sets IV and V.

When looking at the mean errors, LDA seems to give the best performance with a mean error for instance in basis set VI/VII of only 0.17 pm. This value is of the same order of magnitude as was found for CCSD(T). The other xc-potentials are less accurate (~0.6-1.1 pm), but still give results comparable with CCSD in the largest basis set. The worst results are obtained with the BLYP, REVPBE and RPBE potentials, which are the only ones that still give a mean error that is larger than 1 pm in the largest basis set.

The standard deviation in the mean errors is given in Table 4.2.3, which show the same improving trend on increasing the basis set size.

TABLE 4.2.3 STANDARD DEVIATIONS (pm)

XC-potential	I	II	III	IV	V	VI	VII
frozen core							
Becke-Perdew	3.47	3.80	0.94	0.61	0.59	-	-
BLAP3	3.51	4.64	1.64	1.35	1.13	-	-
BLYP	3.55	4.32	1.38	1.09	0.89	-	-
LDA	3.59	2.85	1.28	1.21	1.46	-	-
PBE	3.55	3.75	0.89	0.56	0.60	-	-
PW91	3.45	3.76	0.92	0.57	0.56	-	-
REVPBE	3.49	4.03	0.97	0.61	0.53	-	-
RPBE	3.35	4.07	1.00	0.65	0.52	-	-
average	*3.50*	*3.90*	*1.13*	*0.83*	*0.79*	-	-
all electron							
Becke-Perdew	3.43	3.92	1.00	0.63	0.58	0.59	0.59
BLAP3	3.48	4.74	1.86	1.45	1.21	1.06	1.09
BLYP	3.53	4.40	1.55	1.16	0.95	0.83	0.85
LDA	3.56	2.94	1.04	1.17	1.39	1.53	1.50
PBE	3.55	3.88	0.92	0.63	0.57	0.61	0.61
PW91	3.44	3.87	0.96	0.62	0.57	0.57	0.58
REVPBE	3.33	4.15	1.11	0.67	0.55	0.53	0.53
RPBE	3.39	4.22	1.16	0.71	0.55	0.51	0.53
average	*3.46*	*4.02*	*1.20*	*0.88*	*0.80*	*0.78*	*0.79*

Also here do the frozen core and all electron basis sets perform equally well. There are some small differences, but not as large as the ones observed between basis sets of different size; i.e., whereas the former differences are of the order of 0.1 pm (comparing for instance Becke-Perdew in basis set IV), the latter are a few times larger (comparing for instance the frozen core Becke-Perdew results in basis set III and IV respectively).

LDA no longer performs best in this respect; although it had the smallest mean error of all potentials, the standard deviation belonging to it is the largest (~1.5 pm in basis sets V-VII) of all xc-potentials in the larger basis sets; in the medium basis sets, also a large standard deviation is found for the BLAP3 potential. In fact, for the larger basis sets, all potentials give more or less equal standard deviations except three: BLAP3, BLYP and LDA. As BLYP was also one of the worst potentials in the case of mean errors, it seems that it can not be used with great confidence for obtaining accurate geometries.

The mean absolute errors are given in Table 4.2.4. Just like the mean errors and the standard deviations, they are shown to improve gradually with increasing basis set size.

TABLE 4.2.4 MEAN ABSOLUTE ERRORS (pm)

XC-potential	I	II	III	IV	V	VI	VII
frozen core							
Becke-Perdew	5.68	4.00	1.51	1.24	0.93	-	-
BLAP3	6.27	4.17	1.64	1.38	1.03	-	-
BLYP	6.13	4.32	1.82	1.55	1.21	-	-
LDA	5.05	3.28	1.21	1.14	1.28	-	-
PBE	5.61	3.97	1.45	1.20	0.92	-	-
PW91	5.57	3.82	1.34	1.09	0.80	-	-
REVPBE	5.87	4.28	1.68	1.45	1.13	-	-
RPBE	5.78	4.36	1.76	1.54	1.21	-	-
average	*5.75*	*4.03*	*1.55*	*1.32*	*1.06*	-	-
all electron							
Becke-Perdew	5.64	3.98	1.45	1.30	1.00	0.88	0.89
BLAP3	6.20	4.13	1.59	1.44	1.09	0.93	0.93
BLYP	6.06	4.26	1.76	1.61	1.26	1.09	1.09
LDA	4.97	3.20	1.00	1.11	1.24	1.31	1.29
PBE	5.55	3.95	1.38	1.29	0.97	0.86	0.86
PW91	5.49	3.79	1.27	1.15	0.87	0.76	0.75
REVPBE	5.68	4.25	1.66	1.52	1.20	1.05	1.05
RPBE	5.78	4.35	1.74	1.61	1.28	1.12	1.12
average	*5.67*	*3.99*	*1.48*	*1.38*	*1.11*	*1.00*	*1.00*

Again there is hardly any difference between the results from either a frozen core or an all electron basis set; the difference between the averages of the frozen core and the all electron in basis set IV is for instance 0.06 pm, while the difference between the averages between basis set III and IV for the frozen core is 0.23 pm. The results of basis set V are still improved on going to VI, but increasing the basis set even more (VII) doesn't improve them any more. So, where the accuracy of geometries is concerned, the basis set limit is reached already in basis set VI.

The Becke-Perdew, PBE and PW91 are found to be the best xc-potentials, as they are the only ones that are consistently giving better results than the average for all basis sets, both in the frozen core and the all electron basis sets. The mean error in the basis set limit of 0.89 (Becke-Perdew), 0.86 (PBE) and 0.75 (PW91) pm, are comparable or slightly better than CCSD results (0.89 pm) in the largest basis set used by Helgaker et al. (cc-pVQZ). Although LDA

was performing best for the mean error, in this case it is performing the worst with an average error of ~1.3 pm in the larger basis sets. In fact, for the all electron basis sets, LDA does not improve upon increasing the basis set size after basis set III.

The maximum errors for the xc-potentials in a certain (frozen core/all electron) basis set are given in Table 4.2.5. These results don't show the gradual improvement as the basis set is increased; for instance, going from basis I to basis II the average maximum error increases from 10.7 to 13.4 pm. However, after increasing the basis set even more, it decreases again gradually, except for LDA that exhibits a oscillatory pattern. As the difference between the VI and VII basis sets is small, also these results show that the basis set limit has been reached.

TABLE 4.2.5 MAXIMUM ERRORS (pm)

XC-potential	I	II	III	IV	V	VI	VII
frozen core							
Becke-Perdew	10.52	13.16	3.35	2.51	1.81	-	-
BLAP3	11.96	15.61	6.15	5.44	4.68	-	-
BLYP	11.32	14.91	5.43	4.73	3.95	-	-
LDA	9.88	9.69	3.30	2.32	2.85	-	-
PBE	10.42	12.97	3.31	2.11	1.85	-	-
PW91	10.43	12.83	3.23	2.13	1.64	-	-
REVPBE	10.60	13.90	3.89	2.96	2.11	-	-
RPBE	10.58	14.19	4.21	3.25	2.40	-	-
average	*10.71*	*13.41*	*4.11*	*3.18*	*2.66*	-	-
all electron							
Becke-Perdew	10.42	13.42	3.98	2.68	1.82	1.75	1.75
BLAP3	11.82	15.88	6.72	5.36	4.66	4.10	4.18
BLYP	11.19	15.14	5.91	4.66	3.93	3.38	3.44
LDA	9.73	9.89	2.73	2.29	2.26	2.75	2.72
PBE	10.32	13.25	3.56	2.57	1.86	1.80	1.79
PW91	10.31	13.12	3.54	2.41	1.75	1.59	1.57
REVPBE	10.51	14.23	4.69	3.31	2.37	1.89	1.87
RPBE	10.73	14.53	5.03	3.63	2.49	2.00	2.22
average	*10.63*	*13.68*	*4.52*	*3.36*	*2.64*	*2.41*	*2.44*

The difference between the frozen core and all electron basis sets is not negligible for the maximum error, at least not for the small and medium sized basis sets. For basis set V, the difference is negligible again for some potentials (like Becke-Perdew, BLAP3 or PBE), while for others (LDA, REVPBE) there exists a large difference.

Just like found for the mean absolute errors, there are three potentials (Becke-Perdew, PBE and PW91) that are consistently giving better results than the average for a certain basis set. For these three, the maximum error in the basis set limit (1.75, 1.79 and 1.57 pm respectively) is comparable to MP2 (1.67 pm) and MP4 (1.48 pm) in the largest basis set studied by Helgaker et al., and is substantially better than the CCSD method in the same basis (3.07 pm).

Comparison with Dunning's basis set and the B3LYP potential

The results obtained with Dunning's basis sets for either the BLYP or the B3LYP potential are given in Table 4.2.6. Comparing the BLYP results in the cc-pVDZ basis set (e.g. [3S2P1D] for carbon) with its ADF counterpart (basis III, [4S2P1D]), the mean absolute error of cc-pVDZ is slightly larger than basis set III. However, the maximum error is much smaller for cc-pVDZ than for basis III. The same inconsistencies show up for the mean error and its standard deviation; although the mean error is more or less comparable (1.76 vs. 1.81 pm), the standard deviations are quite different (1.55 vs. 1.06 pm). Therefore, as BLYP does not perform equally well in the cc-pVDZ and the ADF-III basis set, also the B3LYP results (in the cc-pVDZ basis) can not directly compared with the *pure* xc-potentials in basis III.

The size of the cc-pVTZ basis set ([4S3P2D1F] for carbon) is more or less inbetween that of basis set V ([5S3P1D1F]) and VI ([6S4P2D1F]) of ADF. This is reflected in the mean absolute error of the BLYP potential in the cc-pVTZ basis set (1.14 pm), which is somewhere inbetween the results in the V (1.26 pm) and VI (1.09 pm) basis set. Also the maximum error shows this trend: 3.55 (cc-pVTZ), 3.93 (V), 3.38 (VI) pm. There is however a difference for the mean error, which is 0.98 pm in the cc-pVTZ basis, but slightly larger in the ADF basis sets (1.26 and 1.09 pm for respectively basis V and VI); the standard deviation on the other hand is slightly lower in the ADF basis sets: 0.95 (V) and 0.83 (VI) pm vs. 0.99 pm (cc-pVTZ). However, roughly speaking, the cc-pVTZ results for BLYP are similar to those in the V/VI basis sets, which enables a comparison of B3LYP with the *pure* DFT potentials in ADF.

The mean error of B3LYP in the cc-pVTZ basis is almost zero (0.05 pm); however, the standard deviation is almost twice as large as those of Becke-Perdew or PW91. The mean absolute error is either comparable (PW91) or slightly better (Becke-Perdew, PBE) for the B3LYP potential. Finally, the maximum error is slightly larger for B3LYP compared with either Becke-Perdew, PBE or PW91. As a whole, B3LYP performs equally well as the most accurate *pure* DFT potentials (Becke-Perdew, PBE, PW91), although with a slightly increased maximum error and standard deviation of the mean error.

TABLE 4.2.6 BLYP AND B3LYP RESULTS (pm) IN DUNNING'S BASIS SETS

	BLYP cc-pVDZ	BLYP cc-pVDZ	BLYP cc-pVDZ	BLYP cc-pVDZ
mean error	1.81	0.98	0.66	0.05
standard deviation	1.06	0.99	0.87	0.94
mean absolute error	1.90	1.14	0.96	0.73
maximum error	3.80	3.55	2.06	1.93

Metallocenes

Metallocenes are molecules where a metal atom is sandwiched between two cyclopentadiene rings, which can exist in two conformations: *staggered* or *eclipsed* (see Figure 4.2.1).

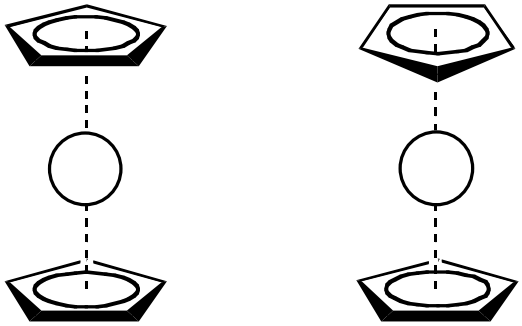


FIGURE 4.2.1. STRUCTURE OF METALLOCENES

The best known example of these metallocenes is ferrocene, which has been studied in great detail in the past by theoretical methods, and showed to be a difficult molecule for which an accurate prediction of the metal-ring distance could be obtained. Early Hartree-Fock calculations reported a Fe-ring distance of 1.88 Å, which is in poor agreement with the experimental distance of 1.66 Å. This could not be improved by employing larger basis sets, as it was established[232] that the Hartree-Fock limit is only slightly better than the early HF calculation (1.872 Å). Normally, such a poor performance could be improved upon by using MP2 calculations; for instance, the systematic study by Helgaker et al. (described earlier in this section) showed a mean absolute error of 13.0 pm for Hartree-Fock, and only 2.4 pm for MP2. However, MP2 results on ferrocene[232] show a similar dramatic performance. However, unlike HF that overestimates the Fe-ring distance, MP2 underestimates it at 1.47-1.49 Å[232] (depending on the number of electrons correlated). Calculations employing the CASSCF and CASPT2 method perform much better in this respect with Fe-ring distances of 1.716 and 1.617 Å respectively[233]. Subsequently, correcting the results for Basis Set Superposition Errors, an estimated equilibrium value of 1.643 Å was obtained, which is in good agreement with the experimental value of 1.66 Å. Also CCSD and CCSD(T) calculations predict an equilibrium distance which is in good agreement with the experimental data, with respectively 1.672 Å and 1.660 Å[223]. Note however, that the coupled cluster distances were obtained by performing single point energies at three Fe-ring distances, and then fitting the potential energy curve with a second order polynomial to obtain the equilibrium distance.

The metallocenes studied in this section are manganocene (either doublet or sextet), ferrocene (singlet), cobaltocene (doublet or quartet), nickelocene (singlet or triplet) and ruthenocene (singlet). The geometry data from experimental investigations (taken from ref. 234) are given in Table 4.2.7.

TABLE 4.2.7 EXPERIMENTAL GEOMETRIES METALLOCENES (Å)

metal M	*r(C-H)*	*r(C-C)*	*r(M-ring)*
Mn (*doublet*)	-	1.418	1.720
Mn (*sextet*)	1.125	1.429	2.041
Fe	1.104	1.440	1.661
Co	1.095	1.430	1.722
Ni	1.083	1.430	1.823
Ru	1.130	1.439	1.823

The distance of the metal to the center of the cyclopentadienyl rings differs quite some amount for the five metal atoms as well as the multiplet state of the complex. For instance for the manganocene, a difference of more than 0.32 Å in the distance is observed between the doublet and sextet state.

The geometry of the metallocenes was optimized using all exchange-correlation potentials that had been used already in the first part of this section, employing a few different basis set sizes, ranging from a minimal basis (I) to a triple zeta valence basis plus double polarization functions (VI). The core electrons were either taken explicitly into account (*all electrons*) or frozen in the calculations; in the latter calculations, the frozen core electrons comprise the 1s electrons of carbon and for the first row transition metals up to either 3p (*frozen core 3p*; for Ru up to 4p) or up to the 2p level (*frozen core 2p*; for Ru up to 3d). For all three types, the optimizations were done in either a non-relativistic or a scalar relativistic (ZORA) calculation. The mean absolute errors of the computed distances for the eight xc-potentials in the four basis sets are given in Tables 4.2.8 (non-relativistic) and 4.2.9 (scalar relativistic).

A general improvement of the accuracy is shown by increasing the basis set size, which is most obvious for the *all electron* calculations. The mean absolute error, averaged over the eight xc-potentials, decreases from 7.70 pm to 1.77 pm. However, there are some potentials that do not show this pattern; the error of the LDA potential for instance decreases in going from basis set I (minimal basis) to II (double zeta valence), and increases if one uses larger basis sets. The same pattern had already been observed in the first part of this section, but there it emerged only after basis set III (double zeta valence plus polarization) and the error in basis set IV (triple zeta valence plus polarization) was still smaller than in basis II. Here, the LDA error is larger in basis IV than in basis II, both if one uses the *frozen core* or the *all electron* calculations.

The difference between the *frozen core 3p*, *frozen core 2p* and *all electron* results is not negligible. For the minimal basis (I), the *frozen core 3p* gives a better performance than the *all electron* calculations; for the larger basis sets, the situation is reversed. In basis II, the best performance is observed with the *frozen core 2p* option, which gives a mean absolute error that is 0.2-0.3 pm smaller than that for the *all electron* case. In basis set IV, these two options perform equally well, and considerably better (0.8-0.9 pm) than the *frozen core 3p* option.

TABLE 4.2.8. MEAN ABSOLUTE NR[a] ERRORS (pm) FOR 1ST ROW METALLOCENES

XC-potential	I	II	IV	VI
frozen core 3p				
Becke-Perdew	7.43	2.77	2.25	-
BLAP3	6.88	5.66	4.99	-
BLYP	7.36	4.52	3.92	-
LDA	8.55	1.23	1.55	-
PBE	7.50	2.52	2.02	-
PW91	7.57	2.53	2.12	-
REVPBE	6.58	3.29	2.61	-
RPBE	7.00	3.53	2.81	-
average	*7.36*	*3.26*	*2.78*	-
frozen core 2p				
Becke-Perdew	-	1.77	1.39	-
BLAP3	-	4.58	4.05	-
BLYP	-	3.49	2.99	-
LDA	-	2.01	2.46	-
PBE	-	1.60	1.29	-
PW91	-	1.60	1.33	-
REVPBE	-	2.25	1.72	-
RPBE	-	2.49	1.92	-
average	-	*2.47*	*2.14*	-
all electron				
Becke-Perdew	7.68	2.05	1.39	1.00
BLAP3	7.21	4.92	4.08	3.18
BLYP	7.69	3.78	3.05	2.17
LDA	8.81	1.87	2.47	3.05
PBE	7.83	1.82	1.28	1.11
PW91	7.82	1.79	1.34	1.13
REVPBE	7.30	2.53	1.72	1.18
RPBE	7.23	2.78	1.93	1.30
average	*7.70*	*2.69*	*2.16*	*1.77*

a) NR=non-relativistic results

The "best" exchange-correlation potentials are for the larger basis sets and the *all electron/frozen core 2p* options, Becke-Perdew, PBE and PW91, just like found in the first part of this section. For the *frozen core 3p* option in the larger basis sets, LDA performs best; in basis II, the mean absolute error is even only 1.23 pm, a value not reached by any other potential in any basis set with the *frozen core 3p* option. However, using the *frozen core 2p* or the *all electron* option, this value is reached in the larger basis sets (IV, VI) by other xc-potentials. In the minimal basis, the BLAP3 and REVPBE give generally the "best" performance, but as the mean absolute error is about four times as large as the value in the larger basis sets, it is of limited value.

TABLE 4.2.9. MEAN ABSOLUTE SR[a] ERRORS (pm) FOR 1ST ROW METALLOCENES

XC-potential	I	II	IV	VI
frozen core 3p				
Becke-Perdew	-	2.47	2.01	-
BLAP3	-	5.31	4.71	-
BLYP	-	4.20	3.66	-
LDA	-	1.38	1.77	-
PBE	-	2.22	1.78	-
PW91	-	2.26	1.88	-
REVPBE	-	2.97	2.35	-
RPBE	-	3.22	2.56	-
average	-	*3.00*	*2.59*	-
frozen core 2p				
Becke-Perdew	-	1.54	1.18	-
BLAP3	-	4.21	3.66	-
BLYP	-	3.14	2.63	-
LDA	-	2.31	2.76	-
PBE	-	1.40	1.07	-
PW91	-	1.41	1.11	-
REVPBE	-	1.93	1.44	-
RPBE	-	2.12	1.59	-
average	-	*2.26*	*1.93*	-
all electron				
Becke-Perdew	7.83	1.72	1.19	1.22
BLAP3	7.37	4.53	3.69	2.74
BLYP	7.85	3.43	2.67	1.76
LDA	8.97	2.18	2.80	3.42
PBE	7.92	1.58	1.06	1.38
PW91	7.97	1.59	1.15	1.39
REVPBE	7.40	2.17	1.45	1.08
RPBE	7.33	2.39	1.59	1.12
average	*7.83*	*2.45*	*1.95*	*1.76*

a) SR=scalar-relativistic results using ZORA approach

The influence of the scalar relativistic corrections is, apart from the *all electron* calculations in basis VI, small (0.2 pm) but improving. This effect is of the same order of magnitude as the effect observed due to a slight mismatch between the energy expression and the potential in the ZORA approach, which leads to an optimized geometry with zero gradient that may differ from the point of lowest energy by some 0.1 pm. Therefore, the effect of the relativistic corrections can safely be ignored.

The *all electron* results in basis VI show different patterns for different xc-potentials. For some, like Becke-Perdew, PBE and PW91, the mean absolute error increases relative to the non-relativistic results, while for others, like LDA, the error decreases. Still, these changes are of the same small magnitude (0.2 pm) and therefore not significant.

TABLE 4.2.10. MEAN ABSOLUTE SR/NR[a] ERRORS (pm) FOR RU-RING DISTANCE

XC-potential	II/NR	IV/NR	II/SR	IV/SR	V/SR
frozen core 3p					
Becke-Perdew	9.01	6.69	6.84	4.54	-
BLAP3	15.64	13.43	13.02	10.89	-
BLYP	13.44	11.36	10.90	8.91	-
LDA	3.52	1.60	1.68	0.41	-
PBE	8.18	5.75	6.13	3.63	-
PW91	8.61	6.12	6.18	3.92	-
REVPBE	9.30	6.79	7.20	4.68	-
RPBE	9.72	7.17	7.63	5.04	-
average	*9.68*	*7.36*	*7.45*	*5.25*	-
frozen core 2p					
Becke-Perdew	5.59	3.57	3.99	1.82	-
BLAP3	11.90	9.91	9.89	7.77	-
BLYP	9.84	8.07	7.91	5.99	-
LDA	0.39	1.28	1.02	2.96	-
PBE	4.82	2.73	3.36	0.99	-
PW91	5.12	3.21	3.54	1.32	-
REVPBE	5.87	3.62	4.29	1.91	-
RPBE	6.25	3.97	4.65	2.25	-
average	*6.22*	*4.55*	*4.83*	*3.13*	-
all electron					
Becke-Perdew	-	-	4.82	2.47	1.18
BLAP3	-	-	10.83	8.64	7.42
BLYP	-	-	8.81	6.77	5.52
LDA	-	-	0.27	2.49	3.83
PBE	-	-	4.16	1.64	0.33
PW91	-	-	4.22	2.10	0.60
REVPBE	-	-	5.17	2.62	1.29
RPBE	-	-	5.56	2.99	1.65
average	-	-	*5.48*	*3.72*	*2.73*

a) NR=non-relativistic; SR=scalar-relativistic results using ZORA approach

For the first row transition metals, the effect of the inclusion of relativistic corrections is small, but for ruthenocene, it probably can no longer be safely ignored. As the relativistic corrections will have an effect mainly on the Ru-ring distance and not as much on the C-C or C-H distances, only the Ru-ring distance is taken into account for the mean absolute error. The errors for the xc-potentials in several basis sets and with different options for the core electrons are given in Table 4.2.10. As expected, the relativistic corrections now do have a significant effect on the values of the mean absolute error, which are improved by an amount of 1.4-2.2 pm. The same trends are observed as before, e.g. generally speaking the error decreases with increasing basis set size (apart from the LDA potential), the *frozen core 2p* option performs better than the *frozen core 3p* option (except for the LDA potential), the BLAP3 and BLYP perform significantly less than the other potentials.

The multiplet states of cobaltocene, nickelocene and manganocene lead in all cases to the same ground state, respectively a doublet (cobaltocene), triplet (nickelocene) and doublet (manganocene). The other states are less favored by an amount of 21 kcal/mol (quartet cobaltocene), 14 kcal/mol (singlet nickelocene) and 14 kcal/mol (sextet manganocene) with the Becke-Perdew xc-potential in basis set IV using the *all electron* option. These values differ by a few kcal/mol with different basis sets and/or xc-potentials, but the relative ordering of the multiplet states does not.

The relative ordering of the conformations of the metallocenes is also rather constant over the range of xc-potentials and basis sets; in nearly all cases the eclipsed conformation is favored. The energy difference between the two conformations however differs between the different metallocenes, respectively 0.4 kcal/mol (doublet cobaltocene), 0.1 kcal/mol (quartet cobaltocene), 1.0 kcal/mol (ferrocene), 1.2 kcal/mol (doublet manganocene), 0.01 kcal/mol (sextet manganocene), 0.1 kcal/mol (singlet or triplet nickelocene) and 0.5 kcal/mol (ruthenocene), using the Becke-Perdew xc-potential in basis set IV. For quartet cobaltocene, sextet manganocene and nickelocene, the energy difference is too small to be considered significant, and both conformations are equally favorable. Again, these values change somewhat by using another xc-potential and/or basis set, but the relative ordering of the conformations remains roughly intact.

Conclusions

The geometries of a set of small molecules were optimized using eight different exchange-correlation potential in a few different basis sets of Slater type orbitals, ranging from a minimal basis (I) to a triple zeta valence basis plus double polarization functions (VI). This enables a comparison of the accuracy of the xc-potentials in a certain basis set, which can be related to the accuracies of wavefunction based methods like Hartree-Fock and Coupled Cluster. Four different checks are done on the accuracy by looking at the mean error, standard deviation, mean absolute error and maximum error. It is shown that the mean absolute error decreases with increasing basis set size, and reaches a basis set limit at basis VI. With this basis set, the mean absolute errors of the xc-potentials are of the order of 0.7-1.3 pm, which is comparable to the accuracy obtained with CCSD and MP2/MP3 methods.

In the second part of this section, the geometry of five metallocenes is optimized with the same potentials and basis sets, either in a non-relativistic or a scalar relativistic calculation using the ZORA approach. For the first row transition metal complexes, the relativistic corrections have a negligible effect on the optimized structures, but for ruthenocene they improve the optimized Ru-ring distance by some 1.4-2.2 pm. In the largest basis set used, the absolute mean error is again of the order of 1.0 pm. As the wavefunction based methods either give a poor performance for metallocenes (Hartree-Fock, MP2), or the size of the system makes a treatment with accurate methods like CCSD(T) in a reasonable basis set cumbersome, the good performance of Density Functional Theory calculations for these molecules is very promising. Even more so as DFT is an efficient method that can be used without problems on system sizes of this kind, or larger.

4.3 *Validation of charge analyses*

The concepts of molecular recognition, electron withdrawing/donating groups and electrophilic substitution reactions

Assigning atomic charges in quantum chemical calculations has mainly two purposes: either they should be used in a subsequent classical mechanics calculation, or they should be used as a simple interpretation of the distribution of charge density within the molecule. An important example in both cases is the field of molecular recognition, i.e. how do molecules "feel" the presence of other molecules. The largest contribution to the way molecules "feel" each other is given by electrostatic interactions between the permanent multipole moments of the molecules. Therefore, it is important that the atomic charges give a good representation of the permanent multipole moments of the molecule. Furthermore, the atomic charges should be able to reproduce "chemical intuition", like for instance with the concept of electron withdrawing/donating groups.

In this section, a few standard charge analyses that are available in the ADF program[117,187] are tested on their ability to give a good description for both of the aforementioned purposes. The tested analyses are the Mulliken[235] charge analysis, the Hirshfeld analysis[236], the Voronoi Deformation Density analysis[117] and the Multipole Derived Charge analysis[183] (described in more detail in Section 3.1). The charge analyses are tested on benzene and some benzene derivatives[a], which enables to check the concept of electron withdrawing/donating groups, as well as to check electrophilic substitution reactions.

All molecular multipole moments reported in this section are obtained relative to the center of mass of the molecule.

Molecular recognition

The electrostatic potential in a point r_S due to the charge density $\rho(r_i)$ of a molecule is obtained as:

$$V_{el.st.}(r_S) = \int \frac{\rho(r_i)}{|r_S - r_i|} dr_i \qquad (1)$$

This can be expanded in a molecular multipole expansion within the Buckingham convention[179,180] as:

$$V_{el.st.}(r_S) = \frac{Q}{r} + \frac{\mu \mathbf{r}}{r^3} + \frac{1}{2}\frac{\mathbf{r} \Theta \mathbf{r}}{r^5} + \cdots \qquad (2)$$

with the molecular charge Q, dipole moment μ, quadrupole moment Θ, and distance vector $\mathbf{r}$ with length r. For normal purposes, the expansion can be cut off after the quadrupole

[a] The substituent groups used for the benzene derivatives are: trifluoromethyl, cyanide, aldehyde, chloride, fluoride, methyl, NH_2, NO_2, OCOH, OH, OMethyl, O-minus and SH_2.

moment term, as the contribution to the electrostatic potential is decreasing with increasing multipole level due to the division by r to the power x, where x increases with increasing multipole level.

For all analyses studied here, except the Multipole Derived Charge analysis[183], there is no constraint on the charge analysis to reproduce any of the molecular multipole moments other than the total charge. This leads to rather large absolute deviations between the *expectation value* of the molecular multipole moments of the benzene derivatives, and the multipole moments obtained from either the Mulliken, Hirshfeld or Voronoi charges (see Table 4.3.1). The dipole moments deviate on average some 0.2-0.3 a.u., while it is even worse for the quadrupole moments. The Mulliken analysis is in both cases giving the largest deviation, with even an average absolute deviation for the xx, yy and zz components of the quadrupole moment of 10.3 atomic units.

The same trend is obtained for benzene, where the first non-zero multipole moment is the quadrupole moment. The expectation value for its zz-component is −5.568, which is well reproduced by the Multipole Derived Charge analysis[183] with a value of −5.595 a.u.; the Hirshfeld and Voronoi analyses perform less with values of −2.026 and −2.272 respectively. The Mulliken analysis again performs rather poorly, and even predicts a quadrupole moment with the wrong sign: +8.883 a.u.

TABLE 4.3.1. AVERAGE ABSOLUTE DEVIATIONS (A.U.) FROM THE EXPECTATION VALUE OF MOLECULAR MULTIPOLES OF BENZENE AND BENZENE DERIVATIVES

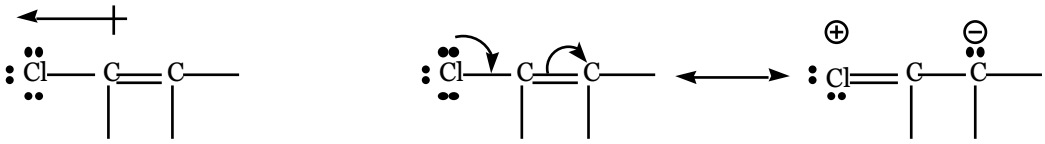
analysis	charge	dipole	quadrupole	quad. xx/yy/zz
Mulliken	0.0	0.30	5.27	10.29
Hirshfeld	0.0	0.26	1.39	2.56
Voronoi	0.0	0.23	1.24	2.28
MDC	0.0	0.03	0.30	0.49

Although also the charges from the Multipole Derived Charge analysis[183] give molecular multipoles that deviate to a small extent from the expectation value of the multipole moments, there is a difference in the reason why these deviations occur. Unlike the other analyses, the molecular multipole moments are *by construction* constrained; however, the analysis uses the moments from the *fitted* density, not from the *"exact"* density (see Section 3.1). There are some small differences between the two sets of multipole moments, which results in the deviations reported in Table 4.3.1.

Electron withdrawing or donating effects

Atoms and functional groups can be organized according to their ability to donate or withdraw electrons either by *inductive* or *resonance* effects[237] (see Figure 4.3.1). For many substituent groups, both effects work in the same direction, but there are some groups (like

the halogen atoms) that exhibit opposing effects for the two cases. These abilities to donate or withdraw electrons have a marked effect on reaction rates for several organic reactions.



Inductive *Resonance*

FIGURE 4.3.1. INDUCTIVE AND RESONANCE SUBSTITUENT EFFECTS

In the last part of this section, a typical organic reaction that shows a large dependence on substituent effects will be considered. The substituent effects found can either be due to (de)stabilization of the ground state or (de)stabilization of the transition state of this reaction. This part of the section only deals with the inductive effects, as these are the ones involved in the ground state properties of the molecules. In the last section, also the resonance effects will be taken into account.

In Table 4.3.2, the substituent effects on the charge distribution is presented as predicted by the four charge analyses. These effects are obtained by taking the difference between the charge of a hydrogen atom in benzene and the total charge of the substituent groups, which are corrected for the change of total charge of the molecule (as in the case of O^- and SH_2^+). A negative value indicates therefore an electron-withdrawing effect, a positive value an electron-donating effect.

TABLE 4.3.2. INDUCTIVE SUBSTITUENT EFFECTS

substituent	*Mulliken*	*Hirshfeld*	*Voronoi*	*MDC*
CF_3	+0.110	−0.091	−0.088	−0.170
CN	+0.014	−0.191	−0.204	−0.273
COH	+0.046	−0.129	−0.122	−0.170
Cl	−0.043	−0.079	−0.124	−0.194
F	−0.453	−0.138	−0.129	−0.285
Me	+0.073	−0.031	−0.010	+0.058
NH_2	+0.095	+0.019	+0.028	−0.245
NO_2	−0.036	−0.198	−0.203	−0.378
OCOH	−0.161	−0.143	−0.156	−0.354
OH	−0.074	−0.063	−0.044	−0.204
OMe	−0.057	−0.047	−0.042	−0.240
O^-	+0.421	+0.485	+0.432	+0.140
SH_2^+	−0.313	−0.389	−0.429	−0.520

Experimentally[237], electron-donating inductive effects are found for O- and the methyl group, while the other substituents are found to have an electron-withdrawing effect.

The Mulliken analysis does not perform completely reliable in this respect. It predicts for instance electron-donating effects for CF_3, CN, COH, NH_2, while this should be a withdrawing effect. However, for the other groups, the correct effect is predicted. The Hirshfeld and Voronoi analyses perform equally here, in that they both predict the correct effects except for two cases: methyl and NH2. For the first a donating effect is predicted and for the second a withdrawing, while in both cases it should be opposite. Generally speaking though, they perform better than the Mulliken analysis, but not as good as the Multipole Derived Charge analysis[183]. This analysis gives a good prediction for the substituent effects for all cases, including the cases where the other analyses fail.

Electrophilic substitution reactions

In electrophilic aromatic substitution reactions, substituent groups have two effects. They have an effect on the reaction rate as well as on the position of the substitution. This is because the reaction may proceed at either the *ortho*, *meta* or *para* position (see Figure 4.3.2).

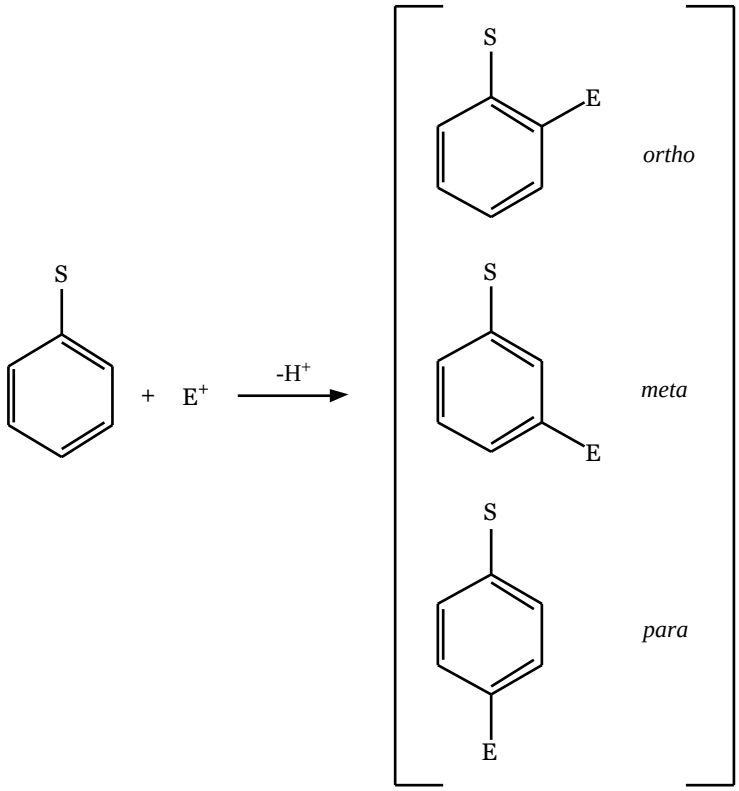


FIGURE 4.3.2. ELECTROPHILIC AROMATIC SUBSTITUTION REACTION

Upon bonding of the electrophile E^+ to the benzene ring, a cyclohexadienyl cation is formed that is stabilized by resonance (as indicated for the *ortho* position in Figure 4.3.3). For both the *ortho* and *para* positions, a resonance structure is available where the positive charge is adjacent to the substituent (Figure 4.3.3, structure on the right), while the *meta* intermediate does not have such a structure. Therefore, if the substituent is electron-donating through resonance, it will stabilize the cation intermediate and enhance the

reaction. Because of the delocalization of positive charge between the ring and substituent in the case of *ortho* or *para* addition, this substituent leads primarily to *ortho* and *para* substitution products. On the other hand, if the substituent is electron-withdrawing through resonance, the reaction will proceed slower and the product will be primarily *meta* oriented.

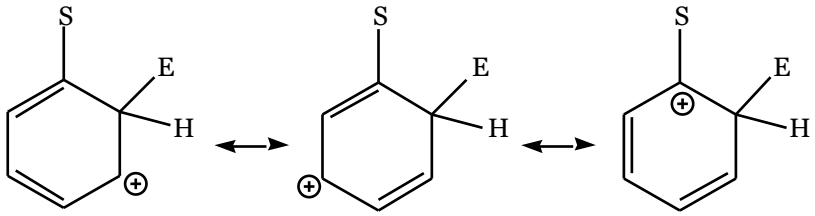


FIGURE 4.3.3. INTERMEDIATE CATION FORMED

An exception to these simple pictures is given by the halogen atoms. Chlorobenzene undergoes nitration about 1/30 as fast as benzene, but the orientation is ortho-para. Here, the difference between *inductive* and *resonance* electron withdrawing/donating effects play a role. The *inductive* electron-withdrawing effect slows down the reaction, yet the *resonance* electron-donating effect favors the *ortho*/*para* oriented products.

TABLE 4.3.3. CHANGE IN CHARGE OF *O*, *M* AND *P*-POSITIONED ATOMS (0.01 A.U.)

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	*m*
Mulliken													
o	5.0	2.6	3.2	-1.6	-6.1	2.3	-3.5	2.8	-1.1	-6.3	-6.4	-11.1	1.5
m	-0.6	-0.4	-0.9	-0.4	0.1	-0.8	-0.8	-0.7	-0.4	-0.6	-0.9	-3.3	2.9
p	1.6	2.3	2.1	0.5	-0.8	-0.4	-2.9	2.5	-0.1	-1.6	-1.8	-7.2	3.3
Hirshfeld													
o	0.7	1.8	1.7	-1.0	-1.5	-0.5	-2.7	1.3	-0.6	-2.2	-2.3	-7.1	1.7
m	0.9	1.1	0.7	0.5	0.6	-0.2	-0.5	1.3	0.6	0.1	-0.2	-5.1	4.4
p	1.3	1.5	1.6	-0.1	-0.6	-0.6	-2.7	2.0	0.1	-1.6	-1.7	-9.0	4.5
Voronoi													
o	1.3	1.7	1.8	-0.5	-2.3	-0.6	-3.0	1.6	-0.7	-2.9	-2.8	-5.7	1.4
m	0.8	0.8	0.4	0.6	0.9	0.1	0.3	0.8	0.7	0.4	0.4	-2.7	3.2
p	1.2	1.4	1.7	0.0	-0.3	-0.3	-2.3	1.7	0.1	-1.1	-1.5	-7.1	3.3
MDC													
o	0.0	2.2	0.9	0.8	-3.0	3.2	-6.1	-0.2	-3.5	-3.7	1.8	-10.0	2.6
m	0.4	-0.7	-0.5	-0.3	-0.8	-1.1	-0.9	-1.2	-1.0	-0.6	-0.8	-6.3	3.4
p	1.1	1.9	3.0	0.2	-1.6	-2.6	-4.8	2.4	-0.3	-3.1	-3.3	-9.3	3.6

a: CF_3, *b*: CN, *c*: COH, *d*: CL, *e*: F, *f*: Me, *g*: NH_2, *h*: NO_2, *i*: OCOH, *j*: OH, *k*: OMe, *l*: O^-, *m*: SH_2^+

Here we check the ability of the charge analyses to predict the orientational substituent effect by checking the charges on the carbon atoms at the *ortho, meta* or *para* positions. In Table 4.3.3, the change of charge for the carbon atoms in the *ortho*, *meta* and *para* positions is given relative to the charge of carbon in benzene. As the reaction involves a positively charged electrophile, either of the *ortho*, *meta* or *para* positions will be favored if the charge on the carbon atom will be more negative, and unfavored if it is less negative. These values can not be used in a strict sense, as both the reaction rate and orientation preference depend on the energies of the transition states leading to the cation intermediate, and not on the charges on the atoms. However, some trends can be observed for the resonance effects. The *resonance* electron-withdrawing effects[237] of NO_2, CN and COH is well predicted by all charge analyses, leading to a favored *meta* oriented product. Also the *resonance* electron-donating effects[237] of for instance OH or NH_2 leading to *ortho*/*para* oriented products is predicted well by all analyses.

Conclusions

Four charge analyses (Mulliken, Hirshfeld, Voronoi Deformation Density and Multipole Derived Charges) have been tested on three properties. The first is their ability to reproduce the molecular multipole moments, which is of great importance for molecular recognition. The Multipole Derived Charges give a good representation of the molecular multipoles *by construction*, while the other three analyses perform poorly (Mulliken) or reasonable (Hirshfeld, Voronoi). For the last three methods, the average absolute deviations of the multipole moments is of the order of 0.2-0.3 a.u. for the dipole moment, and 5.3 (Mulliken) or 1.3 (Hirshfeld, Voronoi) a.u. for the quadrupoles.

The second investigated property is the *inductive* electron-withdrawing/donating effect of substituents. The same trend is shown as for the multipole moments: the Mulliken scheme predicts a wrong substituent effect in four cases, Voronoi and Hirshfeld in two cases, while the Multipole Derived Charges predict the correct effect in all cases.

The last property studied is the substituent effect on the *ortho*/*meta*/*para*-orientation of an electrophilic aromatic substitution reaction. However, as the orientation preference is determined completely by the energy of transition states, and not charges of atoms, this can be checked only in a qualitative way. In fact, all four analyses are found to give a good prediction for the orientation of electron-withdrawing (like CN or NO_2) or electron-donating (like OH or NH_2) groups *by resonance*.

chapter five

REACTION MECHANISMS

5

CONTENTS

SUMMARY

This chapter deals with the application of Density Functional Theory to reaction mechanisms involving metal atoms as a vital and necessary prerequisite. The first part concerns an organic reaction where zinc plays an important role in the aminoalcohol/thiol promoted asymmetric addition of dialkylzincs to aldehydes. In the second part, the mechanism of the reaction taking place in the active site of the copper enzyme quercetinase is studied at all stages of the reaction.

5.1 *Aminothiol/alcohol promoted reactions*

Asymmetric addition of dialkylzincs to aldehydes studied by high-level ab initio calculations

For obvious reasons, the area of catalytic enantioselective synthesis has received an overwhelming amount of interest within organic and inorganic chemistry disciplines. One of the most intriguing and promising observations within this area is the so-called non-linear effect (or chiral amplification); i.e. the achievement of an enantiomeric excess (e.e.) of a product exceeding the e.e. of the catalyst used. For instance, exploitation of this effect obviates the need to obtain catalysts with a very high e.e., which usually is a very costly exercise.

One of the most impressive examples of non-linearity in catalytic enantioselective synthesis has been found in the (2S)-3-exo-(dimethylamino)isobornenol (DAIB) catalyzed methylation of benzaldehyde by dimethylzinc (Figure 5.1.1), as studied by Noyori and co-workers[238-244]. It has been demonstrated that this particular reaction exhibits an extremely large non-linear effect; e.e.'s of over 90 % have been achieved for the formed (S)-1-phenylethanol using DAIB in an e.e. of just 20 %.

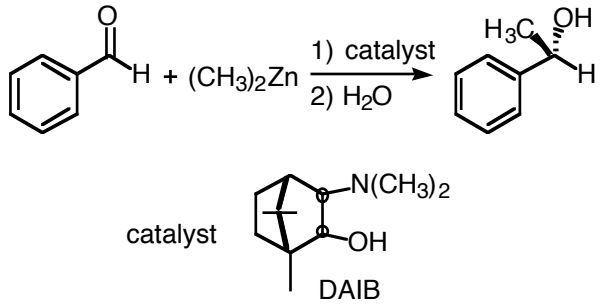


FIGURE 5.1.1. ASYMMETRIC ADDITION OF DIMETHYLZINC TO BENZALDEHYDE CATALYZED BY (DIMETHYLAMINO)ISOBORNENOL

Noyori et al. have studied this reaction in great detail[238-244], both experimentally and theoretically, with the aim to provide more insight into the origin of the non-linear effect. A remarkable and highly relevant experimental observation has been the dramatic decrease in reaction rate when using non-racemic mixtures of both stereoisomers of DAIB instead of using the enantiomerically pure compound as catalyst. This unambiguously indicates that the presence of both stereoisomers causes a partial inhibition of the catalytic cycle. To provide an explanation these findings, an attempt has been made to identify the various intermediate structures of the catalytic cycle by means of quantumchemical calculations at the MP2 level of theory using 2-aminoethanol as model ligand. These efforts have resulted in a proposed catalytic cycle as depicted in Figure 5.1.2.

In this cycle, the methylzinc aminoalkoxide **1** is proposed to be the catalytically active species. Subsequently, an aldehyde and another dimethylzinc moiety are coordinated to methylzinc aminoalkoxide **1** leading to active species **5** via species **3** or **4**, from which the alkylation of the aldehyde takes place. The conformation of **5** is held responsible for the obtained enantioselectivity; on the basis of the aforementioned model calculations. This

view has been corroborated by additional MP2 calculations using DAIB as ligand; both calculations demonstrated that an anti-configuration of the involved transition state is the more stable conformation, thus rationalizing the predominant formation of (S)-1-phenylethanol.

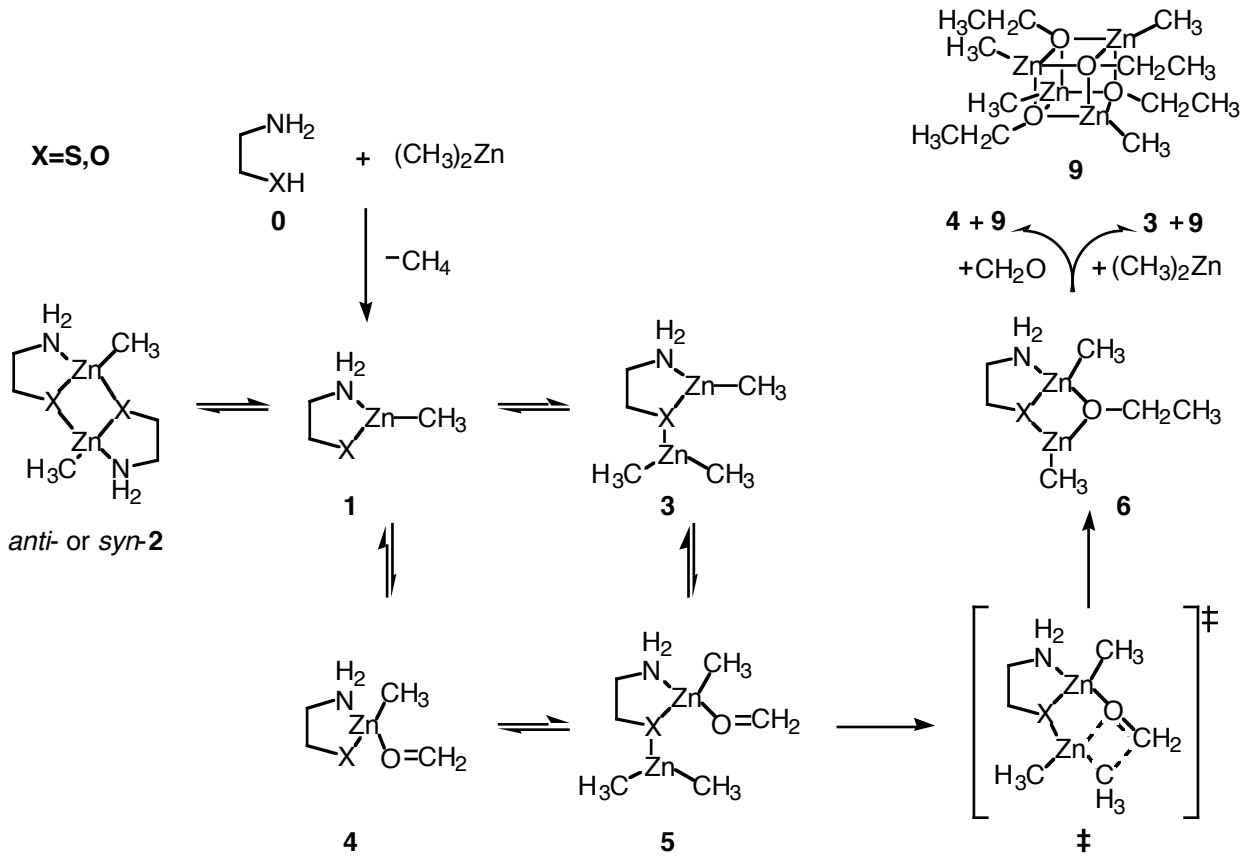


FIGURE 5.1.2. PROPOSED REACTION MECHANISM USED IN MP2 STUDY OF MODEL SYSTEM[243]

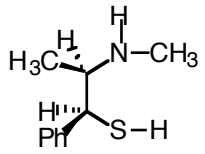
The MP2 model study also confirmed that methylzinc aminoalkoxide **1** is capable of dimerizing to methylzinc aminoalkoxide dimer **2**. Dimer **2** appeared to exhibit two minimal energy conformations; a syn-conformation in which the two methyl groups of the zinc are oriented on the same side of the central Zn-O-Zn-O ring, and an anti-conformation in which these methyl groups are oriented on opposite sides of the central four-membered ring. The MP2 calculations revealed that the anti-conformation of **2** was the more stable of the two by approximately 3 kcal/mole. In addition, the dimers **2** were calculated to be over 35 kcal/mole more stable than two non-interacting monomers **1**, a much larger stabilization in comparison with the other intermediate structures in the catalytic cycle.

The formation of the methylzinc aminoalkoxide dimers **2** has been linked to the occurrence of the non-linear effect of chiral amplification in the DAIB catalyzed alkylation of benzaldehyde. It has been expected that, in the case of chiral monomers **1**, e.g. DAIB, the meso (R,S)-dimer **2** will adopt the anti-conformation, as opposed to the optically active (R,R)- or (S,S)-dimers **2**, which will adopt the syn-conformation. Since the anti-conformer is expected to be the more stable conformer on the basis of the aforementioned MP2 calculations, not only can it be expected that the meso-dimer **2** is predominantly formed, but the release of the (R)- and (S)-monomers **1** from meso dimer **2** is also (much) slower than the

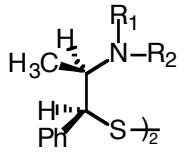
release of the monomers **1** from the optically active syn-dimers **2**. This results in an enhancement of e.e. of free monomer **1** in solution, although the concentration of free monomer **1** is reduced, because a significant amount of monomer **1** is captured in the very stable dimer form **2**, resulting in a negative effect on the reaction rate. Consequently, this mechanistic model provides a highly plausible explanation for both the chiral amplification as well as for the reduction in reaction rate in the case of non-racemic mixtures of DAIB.

Recently, a mathematical study[245] in which experimentally available data has been fitted to appropriate formulas has confirmed the validity of this picture. In addition, it has provided valuable insights into the requirements for the occurrence of chiral amplification in such reactions; it has been suggested that, apart from the required energetic discrimination between the meso- and optically active forms of the aggregates, the release of catalytically active monomer from their aggregated structures should be slow in comparison to the reaction rate of the catalytic process, in order to obtain a substantial chiral amplification in the catalysis. If such release would be fast in comparison to the rate of the catalytic process, both meso and optically active aggregates would be capable of substantial dissociation on the time scales of the catalytic cycle, which would have a detrimental effect on the desired non-linear effect due to the release of both stereoisomers of the catalyst.

The striking success of the DAIB catalyzed alkylation of benzaldehyde has sparked a large number of experimental investigations in an attempt to emulate the results achieved by Noyori et al. One of the more interesting results from these efforts have been presented by Kellogg et al. (see Figure 5.1.3)[246-250].

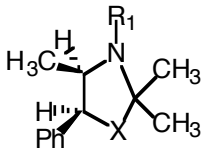


H1



H3

	R_1	R_2
a	H	CH_3
b	CH_3	CH_3
c	CH_3	$CH(CH_3)_2$



H4

	R_1	X
a	CH_3	S
b	H	S
c	CH_3	O
d	H	O

FIGURE 5.1.3. CORRESPONDING AMINOTHIOL CATALYSTS STUDIED BY KELLOGG ET AL.

They demonstrated that the replacement of aminoalcohols by corresponding aminothiol or aminothiol derivatives generally resulted in an increase of the enantiomeric excess of the formed secondary alcohols as well as in an increase of the rate of the reaction. In addition, they proposed a catalytically active species similar to the species proposed by Noyori et al. in

their work. This has led to the assumption that replacement of the oxygen by a sulfur atom in the catalytically active complex improves the catalytic properties of the ligand, an assumption that has been confirmed by subsequent investigations[251-255]. In these investigations, it was shown that the replacement of the oxygen by sulphur increases also the chiral amplification.

This raises an interesting question: why do the aminothiol (derivatives) appear to induce higher enatioselectivity in the formation of optically active secondary alcohols, higher reaction rates as well as a larger chiral amplification ? The work presented in this study aimed at finding an answer to this intriguing question. To obtain this answer, high-level *ab initio* studies have been performed at the Density Functional Theory[1] and MP2 level of theory in which the catalytic cycles associated with the 2-aminoethanethiol and 2-aminoethanol promoted methylation of formaldehyde by dimethylzinc have been compared. In these studies, basis sets of triple zeta quality have been used, and a full counterpoise correction of the BSSE has been included for all single-point energy calculations in order to achieve both qualitatively and quantitatively reliable results. It will be demonstrated that, especially at the MP2 level of theory, large basis set superposition errors can occur, which, if left unaccounted for, can lead to serious misinterpretations of the relative energies associated with the various intermediate minimal energy structures.

Computational details

In this study, the same model reaction as investigated by Yamakawa and Noyori[243] has been used, i.e. the reaction of dimethylzinc and formaldehyde promoted by either 2-aminoethanol or 2-aminoethanethiol (see Figure 5.1.2). As the formation of the cage compound **9** after the formation of compound **6** seems to be favorable and not involved in either the reaction rate or the enantioselectivity, this step in the reaction has been ignored in this study. For each conformation along the reaction path, a full geometry optimization was performed using Density Functional Theory (DFT) calculations[1] with the Becke[120]-Perdew[121] exchange-correlation potential in a triple zeta valence with polarization functions (TZP) basis set of Slater type orbitals.

The geometries of the molecules were obtained using the ADF program[117,187] with the DIIS scheme and the BFGS scheme for updating the Hessian. For the transition state (TS) the Powell update scheme was used, since this is more appropriate for a non-equilibrium geometry optimization. The DFT- calculations were carried out with the ADF program[117,187] on a Linux-cluster of Pentium-nodes as well as on a cluster of IBM/RS6000 workstations. The structures were refined until the gradient was below a threshold of 1.0e-3 a.u. The validity of the obtained structures was checked by inspecting the eigenvalues of the Hessians of these structures, which all proved to exhibit zero negative eigenvalues for the equilibrium structures and one *and only one* for the transition structures, as required by the transition state theory.

The starting structure for locating the transition state was obtained with the SADDLE approach in MOPAC93[219], using the AM1 Hamiltonian. This Hamiltonian, although a bit less accurate than the PM3 option, normally provides transition structures which are very close to those obtained with more robust methods when used in conjunction with the SADDLE approach. Subsequent optimization with *ab initio* methods normally converges rapidly, as

has been the case for our investigations as well. In the here presented studies, we used Cartesian coordinates in MOPAC, but Z-matrix coordinates in the ADF program, in order to be able to use an approximate Hessian in the TS localization. With this choice of coordinates, we were able to explicitly use internal coordinates involved in the transition, which we took for instance to be the two bonds that are broken (Zn-CH3) and formed (CH2-CH3) in the process of going from conformation **5** to **6**. The diagonal element in the Hessian for these two coordinates was given an initial negative value, while the rest of the Hessian was filled using the normal procedure. This scheme proved to converge rapidly.

Using the optimized structures, the interaction energy could be obtained as follows (with $U^{full}(1..N_m)$ the total energy of the system, N_m the total number of monomers and $U^{i,eq}(i)$ the energy of monomer i in its equilibrium geometry):

$$\Delta U_{simple} = U^{full}(1..N_m) - \sum_{i}^{N_m} U^{i,eq}(i) \tag{1}$$

However, this is valid only when using a basis set with an infinite number of functions, or at least a sufficient number to reach the basis set limit. This limit is normally reached only with extensively large basis sets. For smaller basis sets, we can still obtain reasonable interaction energies, after correcting the results for the Basis Set Superposition Error (BSSE) using the full counterpoise method[1]. The BSSE-corrected energy profile is then obtained as:

$$\Delta U_{interaction} = U^{full}(1..N_m) - \sum_{i}^{N_m} \left(U^{full}(i) - U^{i,cmplx}(i) + U^{i,eq}(i) \right) \tag{2}$$

Here, $U^{full}(1...N_m)$ is the total energy of the complex in the complete basis (i.e. of all N_m monomers), $U^{full}(i)$ the energy of monomer i in the complete basis at the complex geometry, $U^{i,cmplx}(i)$ the energy of monomer i in the monomer basis at the complex geometry and $U^{i,eq}(i)$ the energy of monomer i in the monomer basis at its equilibrium geometry. In the following, we will refer to the BSSE correction energies as follows:

$$\Delta U_{correction} = \sum_{i}^{N_m} \left(U^{i,cmplx}(i) - U^{full}(i) \right) \tag{3}$$

The corrected energies ($\Delta U_{interaction}$) are then obtained by taking the sum of ΔU_{simple} and $\Delta U_{correction}$.

The computed Hessians for the structures were used also to obtain thermodynamic properties like the enthalpy H_{rvt}, entropy S_{rvt} and Gibbs free energy G_{rvt} from the rotations, vibrations and translations of the molecules. Once the vibrational frequencies are known, it is straightforward[1,256] to calculate these properties (see Appendix 5.1). The obtained Gibbs free energies G_{rvt} should then be added to the DFT energies to obtain the total Gibbs energies G_{tot}.

The obtained optimized structures were used also to obtain the full counterpoise BSSE-corrected energy profile with single point restricted second-order Møller-Plesset perturbation (MP2) calculations in a 6-311G** basis set for all atoms. All electrons were included in the correlation. These calculations were carried out using the HONDO program[204,225] on a Linux-cluster of Pentium-nodes.

Results and discussion

Summary

The asymmetric addition of dialkylzincs to aldehydes promoted by aminoalcohol or aminothiol has been carefully investigated by high level *ab initio* calculations. The geometries of all relevant conformations in the proposed reaction mechanism (see Figure 5.1.2) were optimized using Density Functional Theory calculations. The energies obtained for these conformations were corrected for Basis Set Superposition Errors, and thermodynamic properties (enthalpy, entropy and Gibbs free energy) obtained from the computed Hessians. These Hessians resulted in zero negative eigenvalues for equilibrium structures and one *and only one* negative eigenvalue for transition structures.

Energies

The energies obtained from the DFT calculations are given in Table 5.1.1. Reported in this Table are the total and the BSSE correction energies, as well as the thermodynamic properties at 298.15 K. For the conformations **6** and **5**-TS no BSSE corrections are needed, as they are obtained with the same number of basis functions as conformation **5**; therefore, the energies can be compared directly. The relative energies resulting from these energies, corrected for BSSE, are given in Table 5.1.2, together with the relative Gibbs free energies G_{rvt} and G_{tot}. The MP2 energies are reasonably similar to the DFT energies (deviations of a few kcal/mol), with a few exceptions. The BSSE corrections are in all cases many times larger than those found in the DFT calculations; in all cases of the same magnitude as the *corrected* interaction energy itself. For instance for the *anti*-**2** (sulphur) conformation, ΔU_{simple} is −18.30 kcal/mol, $\Delta U_{correction}$ 15.94 kcal/mol and therefore the corrected energy $\Delta U_{interaction}$ only −2.37 kcal/mol. The BSSE corrections can therefore play a very significant role and can not be safely ignored if MP2 calculations are used. For the DFT calculations on the other hand, they are much smaller (due to the use of Slater type orbitals) and meaningful interpretations can be obtained also without taking the BSSE corrections explicitly into account.

Structures

Some structural parameters for the geometries of the conformations are given in Table 5.1.3. Some differences between the aminoalcohol and aminothiol geometries are observed (mainly for *Zn-X* distances), but on the whole a large similarity between them is found. On average, the *Zn-S* distances are some 0.4 Å larger than the *Zn-O* distances.

Dimerization

The dimerization of **1** into either an anti- or syn-complex **2** is in all cases favorable by some 10-22 kcal/mol when looking at the DFT energy only. However, the Gibbs free energies are found at much lower values (from −5.8 to +2.2 kcal/mol) and, interestingly enough, are much closer to the experimentally observed ones[254] that were estimated as −2.1 kcal/mol for the thiol and −4.6 for the alcohol. Although the actual aminoalcohol and aminothiol in these studies are different (which may explain the observed differences), the good agreement between experiment and theory may give a clue about the calculated results.

For both the aminoalcohol and aminothiol dimers **2**, the *syn*-conformation is less stable than the *anti*-conformation by some 6 kcal/mol in terms of the DFT energy. For the aminothiol dimers, almost the same energy difference is observed in terms of the Gibbs free energy. The *anti*-conformation of the aminoalcohol dimer **2** however is destabilized with respect to the *syn*-conformation by the entropy contribution to the free energy, leading to a smaller difference in terms of the Gibbs free energy of only 3.5 kcal/mol. This can have a marked effect on the relative stabilities of the optically active (R,R) or (S,S) *syn*-dimers with respect to the meso (R,S) *anti*-dimer, and could explain the higher e.e.'s observed in thiols vs. alcohols.

TABLE 5.1.1. DFT ENERGIES AND THERMODYNAMIC PROPERTIES (AT 298K)

	DFT energy		H_{rvt}	S_{rvt}	G_{rvt}
	Hartree	*Hartree*	*kcal/mol*	*cal/mol/K*	*kcal/mol*
diMeZn	-1.4566		46.50	71.05	25.32
formaldehyde	-0.8051		18.41	53.68	2.41
X=S	*full*	*BSSE*			
1	-2.6564		80.47	96.02	51.84
anti-2	-5.3410	0.0020	162.05	152.74	116.51
syn-2	-5.3317	0.0018	161.77	151.55	116.59
3	-4.1158	0.0007	127.59	129.93	88.85
4	-3.4649	0.0007	99.63	115.70	65.13
anti-5	-4.9260	0.0017	147.39	154.31	101.38
syn-5	-4.9250	0.0017	146.76	149.82	102.09
anti-5-TS	-4.9237		148.50	145.19	105.21
syn-5-TS	-4.9133		147.13	137.57	106.12
6	-4.9958		149.91	139.54	108.30
X=O					
1	-2.7436		81.18	94.27	53.07
anti-2	-5.5254	0.0025	162.55	133.72	122.68
syn-2	-5.5152	0.0023	163.28	145.45	119.91
3	-4.2090	0.0011	128.96	132.43	89.48
4	-3.5543	0.0008	100.98	114.00	66.99
anti-5	-5.0190	0.0021	147.61	142.76	105.05
syn-5	-5.0175	0.0020	149.33	159.84	101.67
trans-5	-5.0203	0.0020	148.80	152.88	103.22
anti-5-TS	-5.0135		148.17	130.49	109.26
syn-5-TS	-5.0062		147.96	131.19	108.84
6	-5.0884		151.45	141.34	109.30

In an attempt to get a reliable comparison for the effectiveness of either aminoalcohol or aminothiol compounds, Kang et al.[255] investigated the catalytic activity of either the pure thiol, or the pure corresponding alcohol, or a 1:1 mixture of the two. It was found that the thiol is approximately eight times more effective than the alcohol. In order to check whether some unexpected results would occur for the *syn*- or *anti*-conformations of such mixed dimers, the same procedure as for the *pure* compounds were followed; i.e. the geometries optimized, energies calculated and corrected for BSSE and thermodynamic properties

obtained from the computed Hessians. The Gibbs free energies for the *anti*- and *syn*-conformations were found to be –4.48 and –0.46 kcal/mol respectively. These numbers are almost equal to the average of the aminoalcohol and aminothiol values for the *anti*- (–4.72 kcal/mol) or *syn*-conformations (-0.09 kcal/mol).

TABLE 5.1.2. RELATIVE ENERGIES (KCAL/MOL) OF CONFORMATIONS

	DFT energy		ΔG_{rvt}	ΔG_{total}
X=S	*full*	*BSSE*		
1	0.00		0.00	0.00
anti-**2**	-16.43	*1.26*	12.83	-3.60
syn-**2**	-10.72	*1.10*	12.91	2.19
3	-3.00	*0.41*	11.69	8.69
4	-2.45	*0.44*	10.88	8.43
anti-**5**	-6.33	*1.08*	21.81	15.48
syn-**5**	-5.64	*1.08*	22.52	16.88
anti-**5**-TS	-4.86		25.64	20.78
syn-**5**-TS	1.69		26.55	28.24
6	-50.12		28.74	-21.38
X=O				
1	0.00		0.00	0.00
anti-**2**	-22.39	*1.57*	16.55	-5.84
syn-**2**	-16.13	*1.47*	13.77	-2.36
3	-6.58	*0.66*	11.09	4.51
4	-3.77	*0.51*	11.51	7.74
anti-**5**	-9.76	*1.29*	24.25	14.49
syn-**5**	-8.83	*1.28*	20.88	12.05
trans-**5**	-10.60	*1.24*	22.43	11.83
anti-**5**-TS	-6.28		28.47	22.19
syn-**5**-TS	-1.74		28.05	26.31
6	-53.37		28.51	-24.86

Complexation of dimethylzinc

The complexation of dimethylzinc to the aminoalcohol/thiol is accompanied by a change of the angle between the two methylgroups in dimethylzinc. While its equilibrium structure is linear, upon complexation the angle decreases to 160° (alcohol) or 165° (thiol conformation). This enables a closer approach to the Zn-complex at either 2.27 (alcohol) or 2.92 Å (thiol), and is accompanied by a favorable DFT energy of –6.6 (alcohol) and –3.0 kcal/mol (thiol). The Gibbs free energy on the other hand is less favorable than conformation **1** by some 4.5 (alcohol) and 8.7 kcal/mol (thiol). If the Zn-complex **1** is visualized as a plane with oxygen/sulphur on the left and nitrogen on the right hand side, dimethylzinc is being found below this plane.

Complexation of formaldehyde

Upon complexation of formaldehyde with conformation **1** to form **4**, a remarkable result is obtained. For the aminoalcohol structure, the formaldehyde molecule is found above the plane of the Zn-complex, while for the aminothiol structure it is found below the plane, like

observed for dimethylzinc. To check whether this may have been influenced by the starting structures for the optimizations, the oxygen and sulphur atoms were switched and the geometries reoptimized; this lead to the same structures that had already been found. The complexation leading to **4** is accompanied by a decrease in DFT energy of –3.8 (alcohol) and –2.5 kcal/mol (thiol), and an increase in Gibbs free energy of 7.7 (alcohol) and 8.4 kcal/mol (thiol).

TABLE 5.1.3. SOME RELEVANT STRUCTURAL PARAMETERS (BONDS IN Å)

compound	*X=S*	*X=O*	*compound*	*X=S*	*X=O*
1					
Zn1-X	2.256	1.870	*Zn1-NH₂*	2.300	2.324
anti-**2**			*syn*-**2**		
Zn1-X1	2.476	2.087	*Zn1-X1*	2.469	2.087
Zn1-NH₂	2.243	2.240	*Zn1-NH₂*	2.277	2.312
Zn1-X2	2.483	2.052	*Zn1-X2*	2.523	2.059
3			**4**		
Zn1-X	2.274	1.902	*Zn1-X*	2.288	1.910[a]
Zn1-NH₂	2.287	2.311	*Zn1-NH₂*	2.260	2.267[a]
Zn3-X	2.921	2.269	*Zn1-O*	2.848	2.719[a]
anti-**5**			*syn*-**5**		
Zn1-X	2.316	1.943	*Zn1-X*	2.316	1.944
Zn1-NH₂	2.237	2.239	*Zn1-NH₂*	2.247	2.254
Zn1-O	2.568	2.524	*Zn1-O*	2.631	2.563
Zn3-X	2.756	2.214	*Zn3-X*	2.769	2.267
Zn3-Me	1.994	2.004	*Zn3-Me*	1.998	2.011
Me-CH₂	4.156	3.939	*Me-CH₂*	4.200	4.026
anti-**5**-TS			*syn*-**5**-TS		
Zn1-X	2.433	2.056	*Zn1-X*	2.427	2.032
Zn1-NH₂	2.226	2.220	*Zn1-NH₂*	2.255	2.284
Zn1-O	2.142	2.139	*Zn1-O*	2.156	2.158
Zn3-X	2.416	2.033	*Zn3-X*	2.426	2.039
Zn3-O	2.532	2.561	*Zn3-O*	2.493	2.471
Zn3-Me	2.179	2.164	*Zn3-Me*	2.220	2.196
Me-CH₂	2.269	2.266	*Me-CH₂*	2.280	2.331
6					
Zn1-X	2.471	2.088	*Zn3-X*	2.432	2.020
Zn1-NH₂	2.287	2.298	*Zn3-O*	1.979	1.967
Zn1-O	2.059	2.041	*O-CH₂*	1.429	1.427

a) Found as *trans*

*Active species **5***

The complexation of both dimethylzinc and formaldehyde may occur in a few different orientations: the formaldehyde may be oriented either *syn* or *anti* with respect to the methyl group connected to the zinc atom of **1**. For the aminoalcohol even a third orientation (*trans*) is observed, where the formaldehyde is located above the plane of **1**, while the dimethylzinc is found a bit below this plane. As the distance between either one of the methyl groups and the CH_2 group of formaldehyde are so far apart, this *trans* orientation is not significantly involved in the reaction leading to conformation **6**. Apart from energetic considerations the

influence of the *syn*/*anti* orientation may be safely ignored for formaldehyde; however for larger aldehydes like the ones used in the actual experiments (benzaldehyde), this *syn*/*anti* orientation determines the enantiomericity of the product. In either the *syn*- or *anti*-conformation, the formaldehyde is already favorably oriented towards either one of the methyl groups of dimethylzinc. For both aminoalcohol and aminothiol, the *anti*-conformation is found to be more stable than the *syn*-conformation in terms of the DFT energy. In terms of the Gibbs free energy this holds also for the aminothiol, but not for the aminoalcohol, which is in both cases to a large extent resulting from entropy contributions. In the transition structures leading from **5** to **6** formaldehyde has moved towards the zinc of **1** by some 0.4-0.5 Å, while also the forming bond between the methyl and CH_2 groups is decreased to a large extent to some 2.27-2.33 Å. At the same time, the distance of the moving methyl group from zinc has increased from 1.99-2.01 to 2.16-2.22 Å, while the C-O distance in formaldehyde increased to 1.28 Å. The transition states are for both the aminoalcohol and aminothiol favorable in the *anti*-orientation by either 4.1 (alcohol) or 7.5 (thiol) kcal/mol. The Gibbs free energy of the transition state is 1.4 kcal/mol lower for the aminothiol promoted reaction than for the aminoalcohol.

Formation of the "product"

The formation of conformation **6** is exothermic after **5** or **5**-TS by almost 21-25 kcal/mol relative to **1**. Compared to the most stable structure of **5**, it is even ~36 kcal/mol more stable. The distance of the formaldehyde oxygen to both zinc atoms is decreased considerably; for both the alcohol and the thiol, the distance towards the zinc of **1** is ~2.05 Å, while it is even closer to the zinc of dimethylzinc at ~1.97 Å. The subsequent steps in the reaction, leading eventually to the product **9** do not involve discriminating features for either reaction rate or enantiomeric excess, and are therefore not investigated.

Conclusions

The aminoalcohol and aminothiol promoted conversion of aldehydes into secondary alcohols by dialkylzinc has been investigated by high-level *ab initio* calculations using 2-aminoethanethiol and 2-aminoethanol as model ligands with dimethylzinc and formaldehyde as reactants. The associated catalytic cycles have been investigated using Basis Set Superposition Error (BSSE) corrected Density Functional Theory (DFT) in a TZP basis set as well as by BSSE corrected MP2/6-311G** calculations. It has been demonstrated that, unlike the DFT method, the MP2 method introduces large BSSE effects in interaction energies of interest, emphasizing the need for the explicit inclusion of BSSE corrections at this level of theory when trying to establish quantitatively accurate estimates of such energies. Finally, it has been demonstrated that the catalytic cycles of the aminoalcohol and aminothiol promoted reactions are quite similar, although the local minima of the intermediate structures of interest on the potential energy surface of the aminothiol promoted reaction are more shallow in comparison to those of the aminoalcohol promoted reaction. This observation is used to provide a rational model in terms of thermodynamic equilibria explaining the enhanced kinetics and chiral amplification in the aminothiol promoted reactions.

Appendix 5.1. Computation[a] of thermodynamical properties[1,256]

Enthalpy H

$$H(T) = H_{translation}(T) + H_{rotation}(T) + \Delta H_{vibrational}(T) + RT$$

$$\begin{aligned} H_{translation}(T) &= \tfrac{3}{2}RT \\ H_{rotation}(T) &= \tfrac{3}{2}RT \quad (RT \text{ for linear molecule}) \\ \Delta H_{vibrational}(T) &= H_{vibrational}(T) - H_{vibrational}(0) \\ \Delta H_{vibrational}(T) &= Nh \sum_{i}^{normal\ modes} \frac{\nu_i}{\left(e^{h\nu_i/kT} - 1\right)} \\ \Delta H_{vibrational}(0) &= E_{zero\text{-}point} = \tfrac{1}{2}h \sum_{i}^{normal\ modes} \nu_i \end{aligned} \tag{4}$$

Entropy S

$$S = S_{translation} + S_{rotation} + S_{vibration} + S_{electronic} - nR\left[\ln\left(nN_{av}\right) - 1\right]$$

$$\begin{aligned} S_{translation} &= nR\left(\tfrac{3}{2} + \ln\left(\frac{2\pi MkT}{h^2}\right)^{3/2}\left(\frac{nRT}{p}\right)\right) \\ S_{rotation} &= nR\left(\tfrac{3}{2} + \ln\left(\frac{\sqrt{u_a u_b u_c}}{s}\right)^{1/2}\right) \\ S_{vibration} &= nR\sum_{i}\left\{\left(\Theta_i e^{\Theta_i} - 1\right)^{-1} - \ln\left(1 - e^{-\Theta_i}\right)\right\} \\ S_{electronic} &= nR\ln w_{electronic} \end{aligned} \tag{5}$$

In these formulas, the following parameters are used:

n	moles of molecules
R	gas constant
N_{av}	Avogadro's constant
M	mass of molecule
k	Boltzmann's constant
T	temperature
h	Planck's constant
p	pressure
I_a, I_b, I_c	principal moments of inertia
u_a,u_b,u_c	$h^2/8\pi I_a kT$, .. , ..
s	symmetry number
ν_i	vibrational frequencies
Θ_i	$h\nu_i/kT$
w_{el}	electronic ground state degeneracy

[a] Formulas given in SI units

5.2 *Quercetinase*

A density functional theory investigation of the reaction mechanism of Aspergillus japonicus Quercetinase[a]

Oxygenases are a class of enzymes that catalyze the incorporation of oxygen atoms from dioxygen into a substrate. They can be divided into two subclasses depending on whether both atoms of dioxygen are incorporated in the substrate (*dioxygenases*) or only one of them (*monooxygenases*). In almost all cases are dioxygenases iron metalloenzymes. Quercetinase (also known as quercetin 2,3-dioxygenase or flavonol 2,4-dioxygenase) is one of a few examples where instead a copper atom is involved[257,258]. This enzyme facilitates the incorporation of dioxygen into quercetin, thereby cleaving its heterocyclic ring to produce the corresponding depside (phenolic carboxylic acid ester) and carbon monoxide (see Figure 5.2.1).

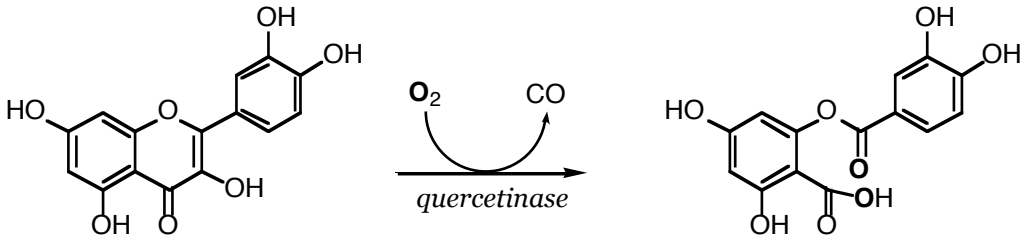


FIGURE 5.2.1. PROCESS CATALYZED BY QUERCETINASE

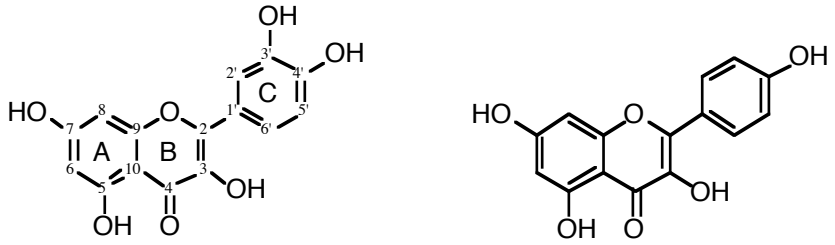
The crystal structure from *Aspergillus japonicus* has recently been solved[259]. The native structure shows an active site with a copper atom coordinated by three histidine residues (His66, His68, His112) at 2.09-2.16 Å and a water molecule at 2.21 Å (*major* conformation); an additional coordination is present, in which a glutamate residue (Glu73) is coordinated at 2.10 Å, while the water molecule is found at a larger distance of 2.41 Å (*minor* conformation). The co-existence of these two conformations is also present in the native EPR spectrum[260], which presents a mixed signal.

Upon binding of the substrates kaempferol or quercetin (KMP and QUE, see Figure 5.2.2)[261,262], a flexible loop (residues 155-169) shields the active site from solvent molecules. The substrate replaces the water molecule as copper ligand, while the glutamate fully coordinates at a distance of 1.97 Å (KMP) in the crystal structure[261,262]. A close interaction between the $O_{\square 2}$ atom of Glu73 and the O_3 atom of the substrate (2.43 Å in KMP) suggests a role for the glutamate residue in the deprotonation of the 3OH group (see Figure 5.2.2 for atom numbering). Deprotonation is likely to occur in order for a proper binding of the substrate to the copper to be established. Although the proton may be exchanged with either solvent or the protein environment, a 1000-fold decrease in activity for the Glu73Gln mutant and the proper positioning of the glutamate residue seems to suggest a role for this residue in the deprotonation[261]. Keeping the proton in close approximity could also be advantageous for protonating the depside product prior to its release.

---

[a] Joint study by R.A. Steiner and M. Swart, who contributed equally to this work.

Another intriguing feature observed in the crystal structure is that the substrates are no longer flat; the C-ring (see Figure 5.2.2 for its definition) is bent out of the plane defined by the rest of the molecule by about 10°. This bending might be the result of steric interactions with the nearby Phe75 and Phe114 residue side chains.



quercetin *kaempferol*

FIGURE 5.2.2. SUBSTRATES INVOLVED WITH QUERCETINASE

A possible reaction mechanism for the quercetinase-mediated dioxygenation of flavonols is shown in Figure 5.2.3. Upon binding of the substrate, accompanied by the closure of the active site through the flexible loop, the Glu73 residue abstracts the proton from the 3OH group of the substrate, to form **2**. A tautomer of this conformation is also possible (**2taut**) where part of the unpaired electron of the copper atom is transferred to carbon 2 giving rise to some radical character on this atom. In the following step, the binding of dioxygen occurs, which can lead to two different routes; in route **a** the (diradical) dioxygen binds to C2 to form the peroxide **3a**, which would lead to the endoperoxide **4** after nucleophilic attack on C4. In the alternative route **b**, the dioxygen binds first to copper (**3b**), a radical combination step occurs (**3b2**) and through a nucleophilic attack, the same endoperoxide **4** is formed. In the next step, the endoperoxide is transformed into the depside **5**, which is accompanied by the leaving of carbonmonoxide. Finally, the depside is protonated (possibly abstracting its proton from the Glu73 residue) and leaves the active site, which is ready for a next cycle.

Route **a** in this proposed reaction mechanism is supported by the crystal structures of quercetinase complexed with QUE and KMP, in which the C2 atom is found pyramidalized. This geometry indicates the possibility of a stabilization of the radical species on C2. The reactive complex arises from an intramolecular radical transfer from the predominant flavonol-Cu(II) species **2**. It is expected to have a low steady-state concentration and to be not easily detectable spectroscopically. To date, EPR evidence for such a species has not been obtained in protic media and only once in the oxygenation reaction of potassium flavonolate to *O*-benzoylsalycilate in DMF[263]. Of similar difficulty appears to be the trapping of endoperoxide **4**, which has not yet been characterized or isolated[263].

In order to obtain insights on the possible reaction intermediates involved in the enzymatic process catalyzed by quercetinase, a Density Functional Theory[1] (DFT) investigation was performed using the experimentally determined copper coordination in the active site. Starting from the crystal structure of quercetinase with the fastest[261] substrate KMP, both routes **a** and **b** have been analyzed as possibilities for dioxygen attack.

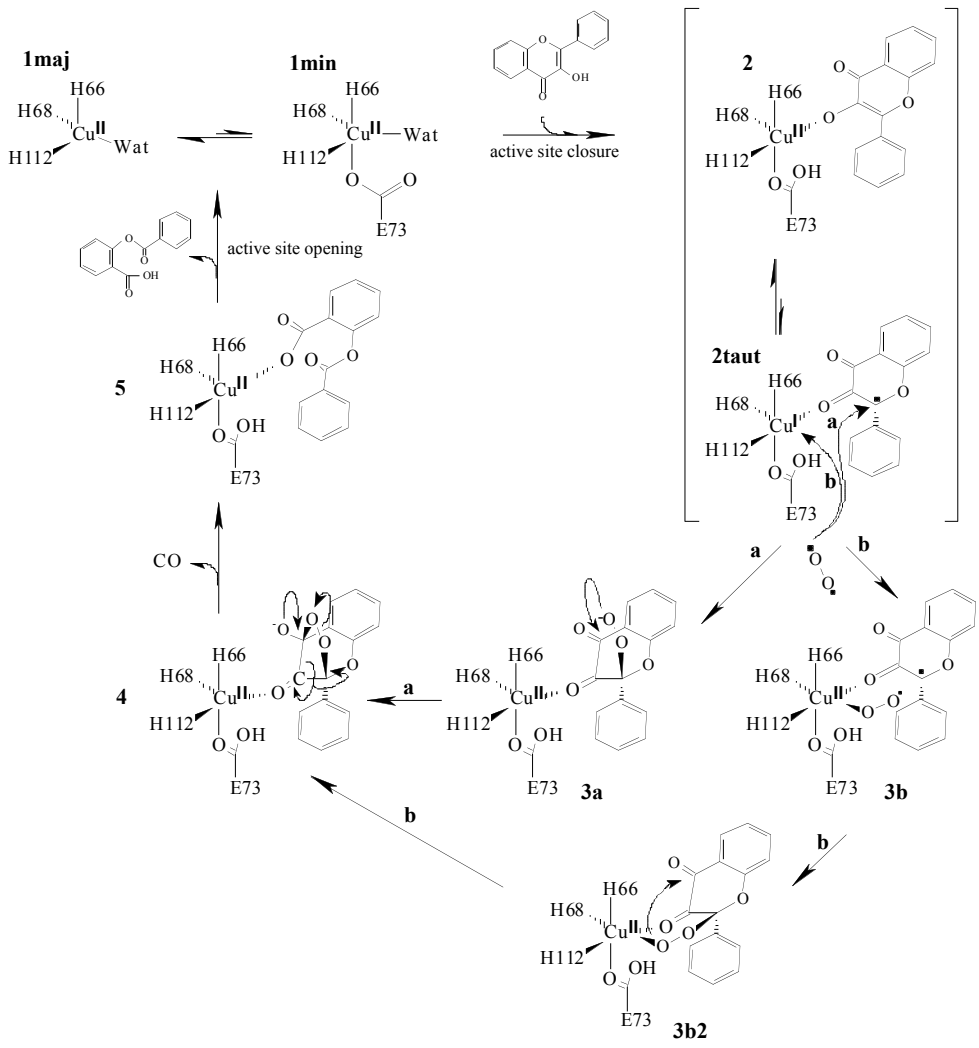


FIGURE 5.2.3. PROPOSED REACTION MECHANISM FOR QUERCETINASE

Computational details

A mechanistic study of the quercetinase-catalyzed dioxygenation of the flavonol kaempferol was carried out using Density Functional Theory[1] (DFT) calculations with the ADF[117,187] program (version 2000.01) using the Becke[120]-Perdew[121] exchange-correlation potential in a triple zeta valence with polarization functions basis set (TZP) of Slater type orbitals. In addition to the copper ion, and other relevant molecules depending on the reaction step under consideration, the proteic copper-ligands (His66, His68, Glu73 and His112), truncated at the Cα atom, were included; on average, this amounted to about 90 atoms. Hydrogens were added where appropriate, resulting in a methyl group at the Cα position. Each structure was subjected to a full geometry optimization, in which the structures are refined until the gradient was below a threshold of $1.0 \cdot 10^{-3}$ a.u.

The recently developed Multipole Derived Charge analysis[183] (see Section 3.1) was used to assign fractional charges and spin densities to the atoms. This analysis uses the atomic multipoles that are obtained by applying an atomic multipole expansion of the Coulomb

potential. The obtained (MDC-q) charges reproduce exactly the atomic multipoles up to the quadrupole moment level and therefore reproduce the Coulomb potential very well, especially in the regions where the higher order moments can be safely ignored. In those regions, the Coulomb potential is reproduced exactly by the MDC-q charges. For systems in a non-singlet ground state (as the ones considered here), the charge density can be obtained for both spins separately. Resulting from these densities are two sets of atomic multipoles (one for spin □, one for spin □), from which two sets of atomic spin charges are obtained. For each atom, the sum of the two spin charges gives the fractional atomic charge, while the difference provides the spin density.

Results and discussion

Complexation of the substrate

The first problem that needs to be solved in order to try to reproduce the experimentally determined E·S state (**2**) lies in the correct positioning of the protons belonging to the system. Owing to the fact that the crystallographic structure used as starting model in the DFT calculations does not contain any information on the location and/or presence of hydrogen atoms, their positions have to be guessed. Three different models have been used in the calculations: **2-dep**, **2-Glu73H** and **2-KMPH**. In the first model (**2-dep**), which has a total charge of zero, the substrate is deprotonated with the proton not present any longer in the system (due to exchange with either solvent or the protein environment); in the second model (**2-Glu73H**), which has a total charge of +1.0, the substrate is deprotonated with the proton located on the $O_{\square 2}$ oxygen (i.e. the oxygen atom not bound to the copper) of the Glu73 side chain; in the last model (**2-KMPH**), the substrate is protonated and Glu73 deprotonated. Based on the short distance between the $O_{\square 2}$ atom of Glu73 and the O_3 atom of the substrate, the second model (**2-Glu73H**) seems to be the most probable, while the third (**2-KMPH**) is chemically not very likely owing to the fact that flavonols are known to be inert towards dioxygen in their protonated form[264]; it has been included for completeness.

The calculations show that the third model (**2-KMPH**) indeed does not need further consideration; the geometry optimization of the structure indicates that a protonated flavonol is incompatible with the presence of a glutamate residue. The 3OH flavonol proton slowly shifts from the substrate to the $O_{\square 2}$ atom of Glu73, leading ultimately to the second model (**2-Glu73H**). This result substantiates the idea of Glu73-mediated deprotonation of the substrate.

The geometries of the optimized **2-dep** and **2-Glu73H** active sites are given in Figure 5.2.4. The overall agreement of the calculated metal-ligand distances and angles with the experimentally observed ones is, for both models, fairly good. However, the short distance between the $O_{\square 2}$ atom of Glu73 and the O_3 atom of the substrate (2.43 Å) found experimentally could be reasonably reproduced only in the **2-Glu73H** model (2.61 Å), in which these atoms are hydrogen bonded. In the other model (**2-dep**), this distance is extremely long (3.16 Å), most likely due to electrostatic repulsion between the negatively charged oxygen atoms (q(MDC-q)=−0.43 for substrate oxygen, −0.54 for Glu73(□2) oxygen). The **2-dep** model exhibits also an unrealistically long Cu-N□2(His112) distance (2.37 Å vs. the experimental value if 2.07±0.04 Å), which is properly reproduced in the **2-Glu73H** model (2.06 Å). On the other hand, the experimental Cu-O□1(Glu73) distance (1.97±0.13 Å) is

more adequately reproduced in **2-dep** (2.08 Å) than in **2-Glu73H** (2.30 Å), but in view of the rather large standard deviation of the experimental Cu-Oε1(Glu73) distance this latter long distance is believed to be less an indication of an improper representation than the deviation from the Cu-Nε2(His112) distance.

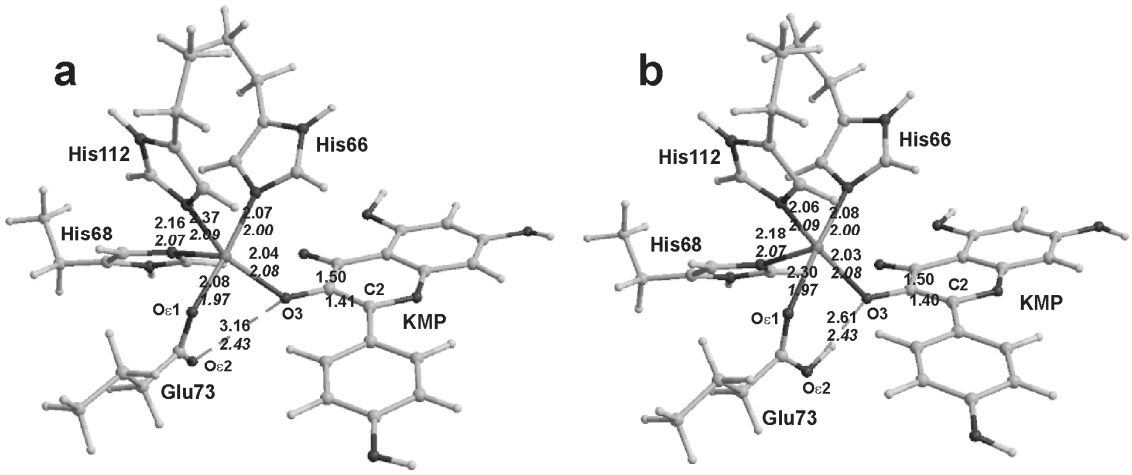


FIGURE 5.2.4. COMPARISON OF COMPUTED AND *EXPERIMENTAL* ACTIVE SITE GEOMETRY OF **2-DEP** (**a**) AND **2-GLU73H** (**b**) MODELS

A direct comparison of the energies of the models favors the **2-Glu73H** model by 290 kcal/mol, but in this comparison a proton (with charge +1) is thought to be transferred infinitely far away. In reality, it will of course have a beneficial interaction with either the solvent or the protein environment, which is ignored in this comparison. A reliable comparison can be made if the "solvation" energies (i.e. the interactions with the protein or solvent environment) are taken into account in the comparison. The "solvation" energies were computed for the **2-dep** and **2-Glu73H** models using the (classical) Direct Reaction Field approach[133-139,141,142,196,197,265-267] (see Section 2.2), where a Boundary Element solution[133,135,137,143] to the Poisson-Boltzmann equation has been used, with a dielectric constant for the protein environment of 20 and using the MDC-q charges for the active site atoms. The proton at infinite distance is assumed to be solvated by water, leading to a solvation energy of –252.39 kcal/mol. The "solvation" energy for the **2-dep** and **2-Glu73H** models are respectively –69.97 and –38.62 kcal/mol, which gives a total stabilization for **2-dep** and $\mathbf{H^+}$ of –322.36 kcal/mol, which is –283.74 more favorable than for the **2-Glu73H** model. Adding the "solvation" and DFT energy differences, the **2-Glu73H** model is more favorable than the **2-dep** model by 6.3 kcal/mol. Therefore, not only on geometric but also on energetic considerations does the **2-Glu73H** model present a better representation for the active site of quercetinase.

An analysis of the spin population in **2-Glu73H** shows that 29 % of the unpaired spin is localized on copper and 28 % on the flavonol ligand (see Table 5.2.1). Significant radical character on the bound substrate molecule is found on the O3 atom (15 %) coordinated to copper and on the C2 atom (13 %). The latter atom is the only one not directly bound to copper having considerable spin density. The partial radical character on C2 supports therefore the existence of a tautomeric complex activated for dioxygen attack (either at C2 or at Cu). An unpaired electron that resides on the O3 atom and to a somewhat lesser extent on the C2 atom

of the substrate is consistent with the results obtained for the oxygenation of potassium flavonolate in DMF, for which a weak EPR signal ascribed to a non-carbon centered radical (g = 2.0038) was observed[263]. Although this latter is believed to proceed with a single electron transfer from the flavonol moiety to dioxygen in order to produce a superoxide anion radical the same **2taut** intermediate is formed in the process.

TABLE 5.2.1. SPIN DENSITIES AND FRACTIONAL CHARGES OF SELECTED ATOMS AND RELATIVE ENERGIES (KCAL/MOL) OF CONFORMATIONS

	2-Glu73H	***3a***	***3b***	***3b2***	***4***	***5***
energy	0.0[a]	−1.20	−2.61	−1.63	+21.15	−87.10[b]
charge						
Cu	+0.53	+0.59	+0.51	+0.52	+0.51	+0.61
C2 (KMP)	+0.20	+0.43	+0.27	+0.37	+0.49	+0.57
C3 (KMP)	+0.22	+0.42	+0.28	+0.29	+0.34	-
O3 (KMP)	−0.58	−0.63	−0.47	−0.45	−0.31	-
C4 (KMP)	+0.29	+0.30	+0.31	+0.31	+0.39	+0.52
O4 (KMP)	−0.43	−0.39	−0.39	−0.39	−0.47	−0.47
Ox1 (O2)	-	−0.22	−0.15	−0.13	−0.24	−0.48
Ox2 (O2)	-	−0.24	−0.14	−0.21	−0.23	−0.52
spin density[c]						
Cu	+0.29	−0.32	−0.27	−0.30	−0.18	−0.35
C2 (KMP)	+0.13	−0.01	+0.03	0.00	−0.03	0.00
C3 (KMP)	+0.02	−0.01	+0.01	0.00	−0.06	-
O3 (KMP)	+0.15	−0.05	+0.01	0.00	−0.01	-
C4 (KMP)	0.00	−0.01	0.00	0.00	+0.01	0.00
O4 (KMP)	+0.03	−0.04	0.00	0.00	−0.21	0.00
Ox1 (O2)	-	−0.04	−0.23	−0.09	−0.01	0.00
Ox2 (O2)	-	−0.05	−0.34	−0.29	−0.06	+0.01

a) energy of isolated dioxygen has been added for relative energy
b) energy of isolated carbonmonoxide has been added for relative energy
c) in **2-Glu73H** the total spin density amounts to one alpha spin in excess, in **3a**/**3b** and further with the presence of triplet dioxygen (two beta spins), the total spin density changes to one beta spin in excess; therefore the sign of the atomic spin densities changes

Dioxygen attack on the C2 flavonol atom

The mechanism of dioxygen attack on an activated ring-position is similar to that proposed for the reaction of the intradiol-type Fe^{3+}-dependent catechol dioxygenases. However, this mechanism has recently been challenged by a DFT study carried out by Funabiki and Yamazaki[268], who reported that they could not obtain a structure of the intermediate arising from dioxygen attack on the catecholate ring of a $[Fe^{3+}(NH_3)_4(catecholate)]^+$ complex. Their explanation was that the catechol ring is positively charged and therefore not prone to dioxygen electrophilic attack; unfortunately, they used Mulliken charges for this explanation, which are known to give the wrong sign for the carbon charges in phenyl rings (see Section 4.3). Moreover, the BLYP[120,227] exchange-correlation potential was used in this

study, which is one of the worst xc-potentials for obtaining geometries, especially when first row transition metals are present (see Section 4.2), and a small basis set was used. The inability to find a structure of the intermediate in their study may be largely influenced by these factors, and therefore does not give any conclusive evidence about the validity of the proposed reaction mechanism.

Even though the C2 atom is positively charged in the **2-Glu73H** model (+0.20), a structure of a dioxygen-substrate adduct **3a** was obtained in this study (see Figure 5.2.5), with an energy that is virtually identical to that of the reactants (see Table 5.2.1). An interesting feature of this complex is that it shows a four-membered 1,2-dioxetane (C3-C2-Ox1-Ox2) structure (Figure 5.2.5). The Ox2 atom of dioxygen is positioned above the C3 atom at a distance of 1.55 Å and not oriented towards the C4 atom as expected in order to form the endoperoxide **4** in the subsequent reaction step. At the time of writing, our various attempts to optimize the structure of peroxide species **3a** in which the Ox2 is near to the C4 atom always failed, leading eventually to the 1,2-dioxetane; a possible peroxide species with Ox2 near C4 is under further investigation.

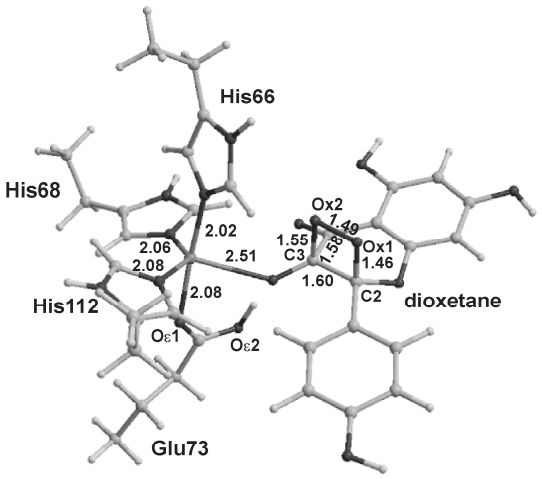


FIGURE 5.2.5. 1,2-DIOXETANE DERIVED FROM DIOXYGEN ATTACK ON KMP

From a geometrical point of view, all protein-metal distances in the 1,2-dioxetane complex are within the range observed experimentally. The Cu-O3 bond is about 0.5 Å longer than found in the experimentally determined E.S complex and 0.2 Å longer than in the computed **2-Glu73H** structure. This longer distance is however still consistent with the dimensions of the active site cavity. Contrary to this, the computed N□2(His68)-Cu-N□2(His112) angle (138°) is outside the normal range observed in the crystal structures (104-115°), and may be an artifact due to the absence of the protein environment in the geometry optimizations.

The existence of a 1,2-dioxetane species as an intermediate in enzymatic reaction should be judged cautiously. Although an early report on the oxygenation of quercetin in aqueous alkaline solution have reported that products consistent with the existence of such an intermediate were obtained[269] this result has later been questioned by Nishinaga and co-workers[270]. Moreover, it is very difficult, as pointed out by the same authors, if not impossible, to reconciliate the existence of a dioxetane intermediate with the experimental

observation from tracer experiment that the substrate's C3 atom is liberated as carbon monoxide, and that none of the atoms of molecular oxygen is incorporated in the leaving CO molecule. The existence of a productive dioxetane intermediate in the enzymatic reactions can, nevertheless, be tested experimentally. These compounds decompose to carbonyl compounds with a very characteristic chemiluminescence[271] which has been observed in the oxygenation of [Cu(phen)(fla)2] (fla = flavonolate, phen = 1,10 phenanthroline) in MeCN[272]. Although no reports which the authors are aware of describe such a phenomenon in quercetinase-mediated dioxygenation of flavonols it is possible that this matter has not been sufficiently investigated. Owing to this result it is clear that some work is necessary to further elucidate this aspect.

Dioxygen attack on copper

Biomimetic studies on a [Cu(II)(fla)(idpa)]ClO4 (fla = flavonolate, idpa = 3,3'-iminobis(N,N-dimethylpropylamine)) complex[273] have indicated the possibility of dioxygen attack on the copper (route **b** of Figure 5.2.3). An analogous proposal has been advanced for the intradiol-type catechol dioxygenases[268], which as already pointed out above, appear to share an activated complex similar to that of quercetinase. The five-coordination of copper in the E S complex indicates that a site for dioxygen binding is, in principle, available at the metal center. The optimized structures arising from such an attack (conformations **3b** and **3b2**) are presented in Figure 5.2.6. From an energetic point of view they are slightly more stable than the reactants, with **3b** and **3b2** being, respectively, 2.6 kcal/mol and 1.6 kcal/mol lower in energy than **2-Glu73H**+dioxygen.

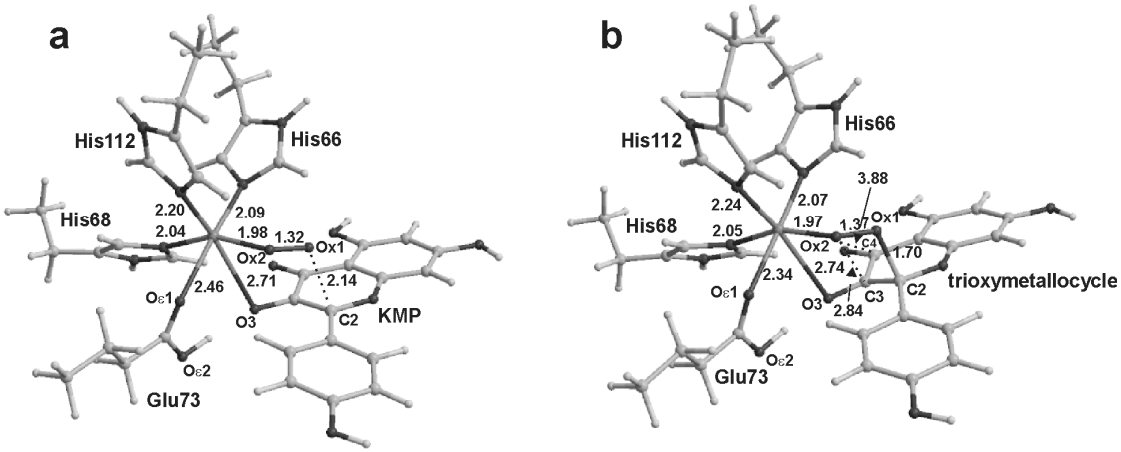


FIGURE 5.2.6. OPTIMIZED STRUCTURES OF CONFORMATIONS **3B** (a) AND **3B2** (b)

The octahedral coordination of structures **3b** and **3b2** features a few long distances. The Cu-O□1(Glu73) distances (2.46 Å and 2.34 Å in 3b and 3b2, respectively) are slightly longer than observed in the octahedral coordination cage of quercetinase in complex with kojic acid (2.26 Å). The Cu-O3(KMP) and Cu-O3(trioxymetallocycle) distances are considerably longer (~0.7 Å) than in the experimentally determined E S complex. However, this result needs confirmation, for example, by including the protein environment in QM/MM calculations.[a]

[a] Including the protein environment in QM/MM has not been tried, as it would increase the already substantial computational cost even more

In **3b** the Cu-dioxygen distance is 1.97 Å and the Cu-Ox2-Ox1 angle is 119.8°, typical for a 'side-on' O2 binding geometry in which, amongst other interactions, an occupied sp_2 hybrid orbital on dioxygen is coordinated to an empty □-symmetry orbital on the metal. An identical value for the Fe-dioxygen distance is found by Funabiki and Yamazaki[268] in their investigation of the mechanism on the intradiol-type catechol dioxygenase. The Ox1-Ox2 distance of 1.32 Å along with the significant spin density population on dioxygen (23 % on Ox1 and 34 % on Ox2) is consistent with its superoxide anion character. In **3b** the Ox1 atom is found at 2.14 Å from C2. The proper positioning and distribution of charges (q(Ox1) = -0.15; q(C2) = +0.27) supports a nucleophilic attack of the superoxide anion on the flavonol to form the trioxymetallocycle **3b2**. In this compound, the Cu-O3-C3-C2-Ox1-Ox2 atoms form a six-membered ring arranged in a 'chair-conformation'. The Cu-Ox2 distance is virtually identical to that in the **3b** complex but the Ox1-Ox2 distance is elongated, which signifies a transition from a superoxide to a peroxide species. However, the trioxometallocycle still appears to maintain considerable superoxide character owing to the rather short Ox1-Ox2 distance (1.37 Å < 1.45 Å) and the considerable spin density at Ox2 (29 %).

According to the proposal advanced by Barhacs and co-workers[273], the endoperoxide **4** might be generated from the complex **3b2** as a result of an intramolecular nucleophilic attack of the Ox2 atom on the C4 carbon. Although such an attack is not impossible, it appears geometrically not favored in the light of the optimized **3b2** structure. Ox2 and C4 are about 3.9 Å apart and the angle between Ox2-C4 and the direction perpendicular to the carbonyl plane is 55°. An attack of Ox2 on the C3 atom, which would form the previously discussed 1,2-dioxetane, seems more favorable. Ox2 and C3 are about 2.75 Å apart and Ox2 forms an angle of 37° with the accessible p-orbital of C3. From an electrostatic perspective, both C3 and C4 atoms are, as expected, apt to nucleophilic attack by Ox2; C3 and C4 have a charge of +0.29 and +0.31 respectively, whereas Ox2 has a charge of −0.21.

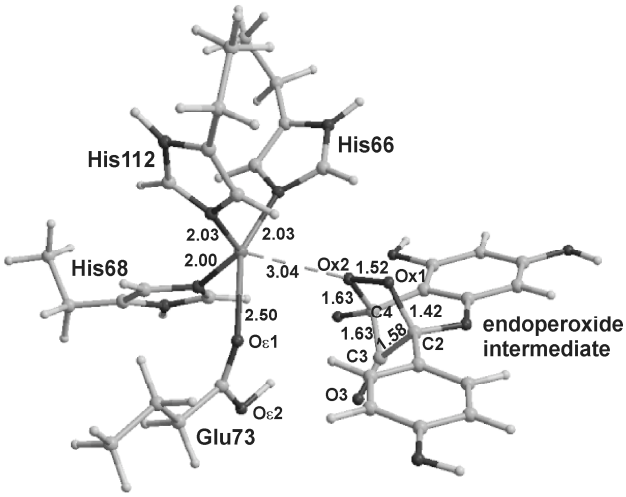


FIGURE 5.2.7. OPTIMIZED STRUCTURE OF ENDOPEROXIDE **4**

*Endoperoxide intermediate **4***

The results discussed above indicate the possibility of the formation of a 1,2-dioxetane species as an intermediate in the quercetinase-catalyzed cleavage of flavonols.

Nevertheless, the five-membered endoperoxide intermediate **4** still represents the best structure in order to rationalize the dioxygenase character of the enzymatic reaction and the liberation of the C3 atom of the substrate as carbon monoxide. The optimization of the structure of the endoperoxide **4** has therefore been performed, with the result presented in Figure 5.2.7. Interestingly, the optimized structure shows that the Ox2-Ox1-C1-C2 dihedral angle has a value of only 9°, which might indicate that the formation of endoperoxide **4** from the 1,2-dioxetane species is not too unfavorable. In the latter structure, the Ox2 and C4 atoms are at 2.4 Å distance. It is therefore possible that a rearrangement of the four-membered ring dioxetane ring driven by nucleophilic attack of Ox2 on the C4 atom produces the five-membered endoperoxide species.

The known instability of such a compound[263] is revealed by its high energy content. It is about 21 kcal/mol less stable than the reactants. Its tendency to easily loose carbonmonoxide with concomitant breakage of the O-O bond is reflected by the very stretched Ox1-Ox2, C2-C3, and C3-C4 bonds. The Ox1-Ox2 distance is 0.07 Å longer than that of a normal peroxide bond. Similarly, the C2-C3 and C3-C4 bonds are considerably longer than in the E S complex (1.58 vs. 1.40 Å, and 1.63 vs. 1.50 Å, respectively). The coordination geometry of this adduct is complex. It can be described as distorted tetrahedral having His66(N□2), His66(N□2), His68(N□2) and Glu73(O□1) being the principal ligands. The O□1(Glu73) atom is further away from the copper (2.50 Å) than in the other structures. As pointed out earlier, it is difficult to assess whether this longer distance is real or an artifact due to the absence of the protein environment in the calculations. In this structure it seems possible that the absence of the protein environment allows the Glu73 residue to shift in order to maintain the O□2(Glu73) atom at hydrogen bond distance to the O3 atom, which moves out of the flavonol plane. Although weak, an interaction might also be present between the copper center and the Ox2 peroxide atom. This appears to be the only direct interaction between the endoperoxide molecule and the catalytic core.

Depside complex ***5***

Air exposure of the anaerobic complex of quercetinase and quercetin produces an EPR spectrum different than that of the native enzyme[260]. This result was interpreted in terms of the formation of an E depside complex. Depside complexation of the quercetinase copper center is not unexpected. The complex [Cu(idpa)(*O*-bs)]ClO4 [idpa = 3,3'-iminobis(N,N-dimethylpropylamine), *O*-bs = O-benzoylsalicylate], which can be obtained from [Cu(idpa)(fla)]ClO4 [fla = flavonolate] with release of CO, has been synthesized and characterized by X-ray crystallography[274]. In this complex the carboxylate group of *O*-bs asymmetrically chelates the copper (Cu-O1 = 1.995(5) Å; Cu-O2 = 2.344(6) Å).

In contrast, the optimized structure of the E depside complex, shown in Figure 5.2.8, reveals that the carboxylate group of the depside product weakly interacts with the copper in a monodentate way (at the long distance of 2.87 Å). This is consistent with the monodentate flavonol binding observed in the anaerobic quercetinase-substrate complexes. From a mechanistic point of view it is interesting to note that in the course of the optimization, the proton originally located on the O□2(Glu73) atom drifts to the Ox2 atom of the depside product causing a progressive elongation of the Cu-Ox2 distance. This result supports the idea that in the last step of the catalytic cycle Glu73 protonates the depside product in order to facilitate its release from the active site.

The line-shapes and g-values of the EPR spectrum of the putative E depside complex resemble those reported for Cu^{2+} spectra of model complexes of trigonal bipyramidal geometry[275,276]. The coordination of the copper site in the E depside complex has therefore been suggested to be also trigonal bipyramidal. The optimized structure shows that although some trigonal bipyramidal character is present (His66-Cu-Glu73 is the axial direction), the site has a predominant square pyramidal character (τ index[275] = 0.40). This contradiction may be explained by the fact that this site shows, similar to the structure of the 1,2-dioxetane discussed before, a large His68(Nε2)-Cu-His112(Nε2) angle (146°). It exceeds by 40° the average experimental value of quercetinase-substrate complexes, and may be the result of the absence of the protein environment in the calculations. Inclusion of the protein environment in QM/MM calculations[a] might result again in a trigonal bipyramidal site as expected on the basis of the EPR experiment.

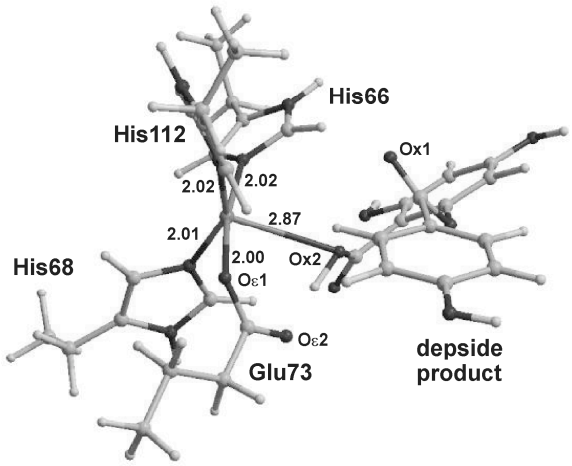


FIGURE 5.2.8. OPTIMIZED STRUCTURE OF E·DEPSIDE COMPLEX **5**

Conclusions

The most generally accepted reaction paths (routes **a** and route **b**) for the quercetinase-mediated dioxygenation of flavonols have been investigated using DFT calculations. The results support the existence of an activated flavonoxy radical-Cu+ complex (**2taut**) that can be attacked by molecular dioxygen. Most likely, the Glu73 copper ligand acts as a base that takes the proton from the incoming flavonol and retains it during catalysis. It is found that the attack of dioxygen on the carbon centered radical of flavonol might lead to a 1,2-dioxetane species. At this moment it is still unclear whether this species really occurs in the catalytic mechanism. Dioxygen attack on the copper center itself appears to be possible also. The trioxymetallocycle **3b2**, which originates from the Cu-superoxide species **3b**, seems to form more easily the dioxetane species than the endoperoxide **4**. This latter intermediate is, however, better suited to explain the chemistry of the enzymatic reaction. As the dioxetane,

[a] Including the protein environment in QM/MM has not been tried, as it would increase the already substantial computational cost even more

3b, and **3b2** species all have virtually the same energy as the reactants, it is not impossible that all these intermediates are unproductive, and that all revert to **2-Glu73H** and free dioxygen, until the proper arrangement for the formation of peroxide **3a** is reached. This elusive peroxide **3a** adduct is under further investigation. However, the five-membered endoperoxide **4** may also be generated by rearrangement of the four-membered dioxetane species. Decomposition of the endoperoxide species **4** is shown to easily form CO with concomitant breakage of the O-O peroxide bond. In the final stage of the reaction Glu73 donates back the proton abstracted in the first step of the mechanism, to facilitate product release.

chapter six

COPPER PROTEINS

azurin

6

CONTENTS

SUMMARY

The creation of a copper force field for use in azurin (and derivatives) is described in the first two sections of Chapter 6, and is followed by sections concerning the axial bonding in the active site of azurin and the validation of the copper force field in Molecular Dynamics of azurin. The creation of the copper force field involves finding suitable atomic charges for the reduced and oxidized state and at stages in-between these states. The effect of the electric field of the charges in the surrounding protein on the energy and atomic charges is checked, and the force constants for copper-ligand interactions determined. In the last section, the vibrational frequencies from either DFT, the copper-ligand force constants or the ones obtained in MD simulations are compared.

6.1 *Redox chemistry*

Charge distribution and energy changes of active sites of copper proteins in redox reactions

In this chapter several copper proteins are investigated by a combined Density Functional Theory (DFT)[1] / Molecular Dynamics (MD)[160] approach. The force field parameters needed for describing the active site in the classical MD simulations are obtained from DFT calculations on the active site, either with or without the rest of the protein present. Density Functional Theory calculations are used since it is an accurate yet computationally cheap method, that enables a treatment of larger systems (in this case the complete active site) at a high computational level. Moreover, as the basic property of DFT is the charge density, it enables to perform calculations on the active site where an electron is removed in steps of 0.2 electron in going from the reduced to the oxidized state. This property of DFT is impossible (or at least very difficult and awkward) to obtain with wavefunction (Hartree-Fock) based methods.

The force field parameters are thus obtained as a function of the removal of the electron, and can be used in MD simulations where an electron is removed in steps from the system. Using thermodynamic integration it is then possible to obtain the redox potential of the copper protein, i.e. the change in free energy in going from the reduced to the oxidized state. This change in free energy A should be obtained after integration over *all* degrees of freedom of the system in a Molecular Dynamics simulation[160], in which the partition function Q_{NVT} is obtained:

$$A = -kT \ln Q_{NVT}$$

$$Q_{NVT} = \sum_i \exp\left(-U_i / kT\right) \tag{1}$$

$$\frac{\partial A}{\partial \lambda} = \frac{-kT}{Q_{NVT}} \frac{\partial Q_{NVT}}{\partial \lambda} = \frac{-kT}{Q_{NVT}} \frac{-1}{kT} \sum_i \frac{\partial U_i}{\partial \lambda} e^{-U_i/kT} = \left\langle \frac{\partial U_i}{\partial \lambda} \right\rangle_{ens}$$

Integrating the derivative of the Helmholtz free energy A with respect to lambda from zero (reduced state) to one (oxidized state), the redox potential is obtained. Experimentally, the redox potential is measured relative to hydrogen, which makes it difficult to compare the absolute values of the redox potentials directly. However, it is possible to compare computed and experimentally found redox potential *differences*.

The copper protein studied is azurin. Although its function in nature is still unclear, it presumably serves as an electron transfer protein. Its structure was solved only in the last twenty-five years[6,7,277,278], and showed three in-plane ligands and two axial bonding groups for the copper ion. The in-plane ligands are two histidines and one cysteine (see Figure 6.1.1). Above the plane resides a methionine; below the plane the backbone oxygen of a glycine is oriented towards the copper. The first four of these residues seem to be common for the type 1 copper proteins, while the fifth group (glycine) is found only in azurin. In Section 6.3, the nature of the interactions between the copper and the axial ligands will be discussed in more detail.

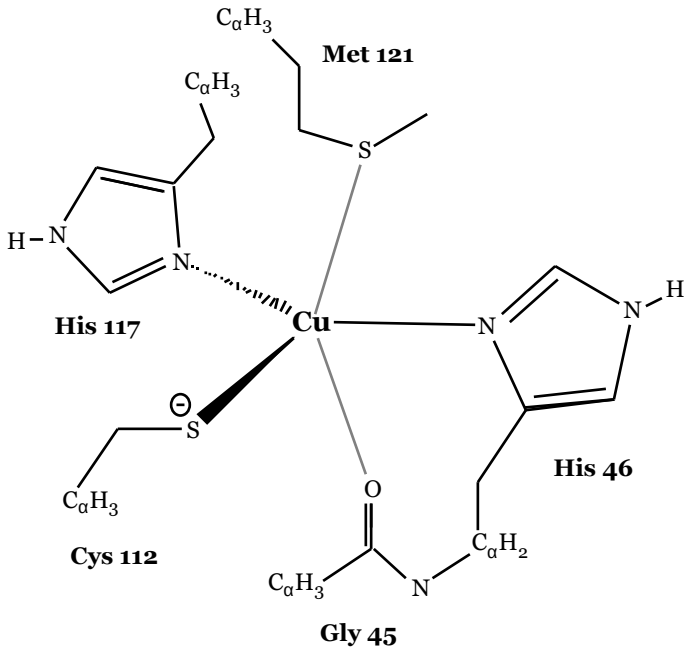


FIGURE 6.1.1. ACTIVE SITE OF AZURIN

In the past, site-directed mutagenesis has been used to study the influence of amino acid residues on the stability and properties of the active site, by specifically replacing residues, both in- and outside the active site. These studies showed a marked dependence of the reduction potential on the type of residue at the 121 position, in a range of 210 to 520 mV[3].

The azurin molecules studied in this section are from *Pseudomonas aeruginosa* (*Pa*) and are *wildtype* at low and high pH (5.5 and 9.0 respectively)[7], the N47D (Asn→Asp) mutant[4] and the F114A (Phe→Ala) mutant[5]. The pH-dependence of *wildtype Pa* azurin has been studied thoroughly[7,72], showing a deprotonation at high pH of the His35 residue, which is located at some 12 Å from the copper ion. This deprotonation seems to have only a negligible effect on the active site geometry[7,72] but does involve a remarkable backbone shift around the His35 position (Pro36-Gly37 peptide flip)[7]. In the calculations in this section the pH can not be taken explicitly into account; it can only be mimicked by using the protein coordinates from the X-ray crystallographic data at either high or low pH, and use an appropriate protonation state for the His35 residue. Although the active site in both cases is the same, the influence of the surrounding protein should be different. The two *wildtype* molecules will be referred to as *wt1* (active site and protein coordinates taken from X-ray data at low pH) and *wt2* (coordinates taken from X-ray data at high pH) from this point on.

The N47D-mutant PDB-file contains a copper ion, although the structure itself was obtained for the zinc derivative, which resembles that of zinc-substituted *wildtype* azurin[59]. Optical and EPR spectra for the copper derivative of the N47D-mutant are similar to *wildtype* and, under the assumption that these should change considerably if the copper ion moves around, this seems to indicate that the structure of the copper derivative is similar to the *wildtype* structure.

Computational details

The Density Functional Theory calculations were performed with the ADF program (versions 2.3.3, 1999.03, 2000.02)[177,178,187] on a cluster of either IBM RS/6000 or Pentium-Linux boxes. In all calculations the Becke[120]-Perdew[121] exchange-correlation potential was used in a triple zeta basis set with polarization functions (TZP). To study the charge distribution in the active site, the Multipole Derived Charge analysis[183] (see Section 3.1) was used, where the MDC-q charges have been used.

The coordinates for the four azurin variants were taken from PDB-files that were obtained from the Brookhaven Protein Database {F114A *Pa* (1azn), N47D *Pa* (1azr), *wildtype Pa* obtained at pH5.5 (4azu; hereafter referred to as *wt1*), *wildtype Pa* obtained at pH9.0 (5azu; *wt2*)}.

Results

The coordinates for the atoms in the active site were taken from the PDB-files, with the amino acid residues cut off at the carbon-alpha position. Hydrogen atoms were added where appropriate and the hydrogen positions optimized while the other positions were kept fixed. This ensured that all experimental data stayed as they were, while only the unknown positions were added. The other positions could have been optimized in our calculations also, but as the rest of protein was not included, the residues would not feel the interactions with the rest of the protein. In Section 6.3 and Chapter 9, this issue will be addressed in more detail.

In all PDB-files, four molecules were found in the asymmetric unit, leading to four subunits, and therefore to four active site structures per azurin variant. Each of these subunits has been treated separately and the results averaged afterwards.

Energies

After optimization of the hydrogen positions, the energies were computed when an electron was removed in steps of 0.2 electron from each active site structure. The energy profile as a function of the removal of the electron is given in Figure 6.1.2 for the four azurin variants. In these plots, the energies are given as a function of the total charge of the system (λ), which corresponds to 0.0 for the reduced state and +1.0 for the oxidized state.

The removal of the electron is accompanied by an increase in energy that fits perfectly to a second degree polynomial. For *wt2* and N47D azurin, a small r.m.s. variation is found for the energies of the four subunits of respectively 1.36 and 1.35 kcal/mol. For *wt1* and F114A azurin, these variations are a bit larger (6.32 and 7.40 respectively). The energy needed for going from the reduced to the oxidized state (ionization potential), is rather similar for the four azurin variants (see *in vacuo* results in first column of Table 6.1.1). The difference between the *wt1* and *wt2* energies is negligible, while the difference with the mutants is still rather small (less than 3 kcal/mol). For comparison, also the ionization potential of an isolated Cu ion is given in Table 6.1.1, which has a much higher value of almost 500 kcal/mol.

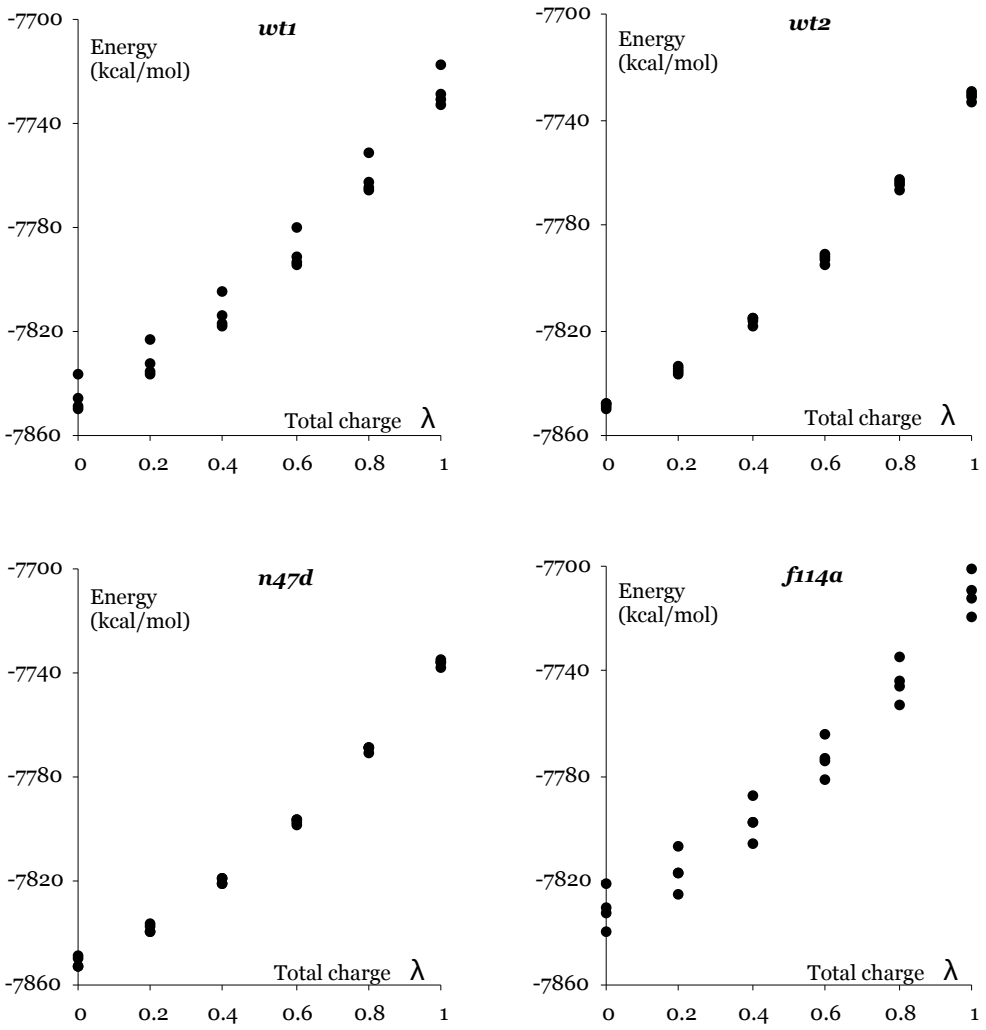


FIGURE 6.1.2. ENERGY PROFILE[a] FOR REMOVAL OF ELECTRON FROM ACTIVE SITE *IN VACUO*

The rest of the protein was taken partly into account in subsequent calculations, in which it was represented by point charges only. The interaction between the active site and the rest of the protein was therefore reduced to electrostatics only. However, as the coordinates of the atoms in the rest of the protein were kept fixed, the only term in the interaction energy (in non-polarizable force fields like AMBER[126] or GROMOS[127]) that changes in going from the reduced to the oxidized state, is the electrostatic energy. Two sets of protein charges were used, either the standard Gromacs[279] charges or the charges obtained with the Multipole Derived Charge analysis[183] for isolated amino acid residues.

The inclusion of the point charges in the calculations has a stabilizing effect on the active sites of all subunits. The energy goes down for all for them, but not by the same amount for all azurin variants and not by the same amount for the reduced or oxidized state. Both result from the total charge contained in the rest of the protein. This is either −2 or −3, depending on the variant and the pH at which the crystal structure was obtained. For *wt1* and N47D azurin it is −2, while it is −3 for *wt2* azurin. F114A has two subunits with a

[a] The four dots at every λ point correspond to the four molecules in the asymmetric unit

protein charge of −2 and two with −3. Due to the negative protein charge, the oxidized state (with a positive charge extra) is stabilized more than the reduced state. This leads to a decrease of the ionization potential (see Table 6.1.1).

TABLE 6.1.1. COMPUTED IONIZATION ENERGIES (KCAL/MOL)

	in vacuo	*GMX surroundings*	*MDC surroundings*
wt1	117.63 ± 0.77	92.88 ± 4.83	71.64 ± 4.79
wt2	117.62 ± 0.76	55.82 ± 4.70	36.25 ± 4.35
n47d	114.83 ± 2.26	48.04 ± 1.67	48.04 ± 0.72
f114a	120.47 ± 2.01	91.68 ± 9.30*	67.24 ± 8.15*
		47.63 ± 0.69#	25.45 ± 2.52#
Cu^+	499.33		

GMX) Gromacs *) Protein charge −2 #) Protein charge -3

Charge distribution

For the active sites *in vacuo* (without the protein environment charges present), the charge distribution has been calculated using the Multipole Derived Charge analysis. Just like the calculated energy, this is also obtained as a function of the removal of an electron from the site. Given in Table 6.1.2 are the residue charges, i.e. the total charge contained in a certain amino acid residue; the distribution of charge within the residues is depicted in Figure 6.1.3 for the reduced state and in Figure 6.1.4 for the oxidized state.

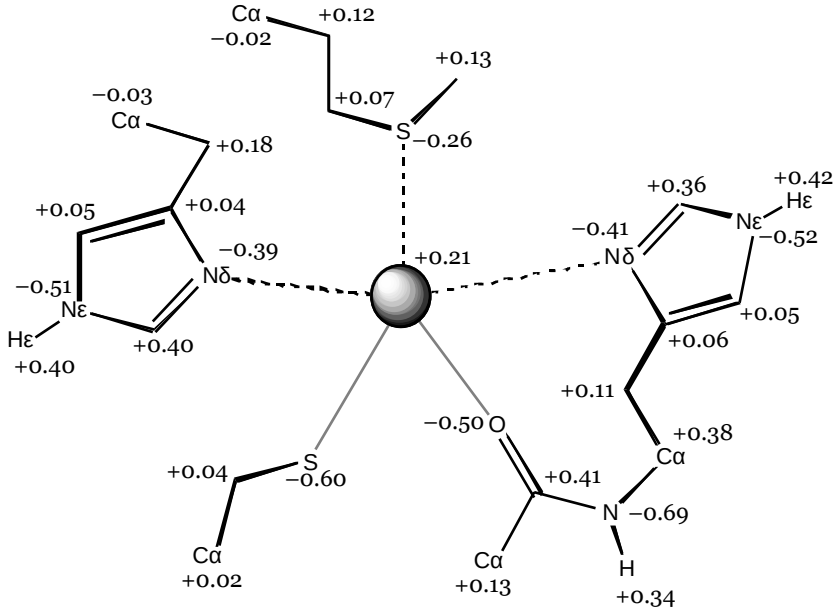


FIGURE 6.1.3. CHARGE DISTRIBUTION IN ACTIVE SITE (REDUCED STATE)

The difference between the four azurin variants is rather small, just like the energies of the active site *in vacuo*. The r.m.s. variation of the residue charges for all 16 active site structures (four subunits of four azurin variants) is rather small with values of ~0.02-0.4 a.u.

TABLE 6.1.2. MDC-Q CHARGES FOR RESIDUES IN ACTIVE SITE

	vacuo	*GMX surroundings*	*MDC surroundings*
reduced state			
Gly45	0.041 ± 0.025	0.030 ± 0.026	0.010 ± 0.025
His46	0.101 ± 0.022	0.166 ± 0.077	0.200 ± 0.059
Cys112	-0.541 ± 0.040	-0.598 ± 0.074	-0.566 ± 0.067
His117	0.141 ± 0.024	0.151 ± 0.032	0.170 ± 0.039
Met121	0.049 ± 0.020	0.019 ± 0.022	-0.023 ± 0.026
Cu	0.209 ± 0.042	0.233 ± 0.040	0.209 ± 0.036
oxidized state			
Gly45	0.060 ± 0.028	0.068 ± 0.025	0.033 ± 0.026
His46	0.239 ± 0.020	0.311 ± 0.074	0.337 ± 0.057
Cys112	-0.030 ± 0.047	-0.088 ± 0.088	-0.047 ± 0.076
His117	0.268 ± 0.017	0.281 ± 0.028	0.307 ± 0.034
Met121	0.135 ± 0.049	0.095 ± 0.028	0.053 ± 0.026
Cu	0.329 ± 0.043	0.334 ± 0.036	0.317 ± 0.034
ox - red			
Gly45	0.019	0.038	0.024
His46	0.138	0.145	0.137
Cys112	0.511	0.510	0.518
His117	0.126	0.130	0.137
Met121	0.086	0.076	0.075
Cu	0.120	0.101	0.108

GMX) Gromacs

In the reduced state, the total charge on the axial groups is small (~0.05 a.u.), and a bit larger on the histidines (0.10-0.14 a.u.); Cys112 has a rather large total negative charge of −0.541 a.u. The copper ion is found not to have its formal charge in either the reduced or oxidized state; instead, its charge is 0.21 in the reduced state and 0.33 in the oxidized state due to back-donation of charge from the ligands. The largest contributor to the back-donation is found to be Cys112 that accounts for more than half of the electron being removed. The other two in-plane ligands account for some 25 % more, while the remainder is coming mostly from Met121 and copper itself; Gly45 accounts only for 2 %. This leads to a total charge on Cys112 around zero, while both histidines have a charge of around 0.24-0.27 a.u. The charge on Gly45 remains small, while some considerable amount of positive charge is found on Met121.

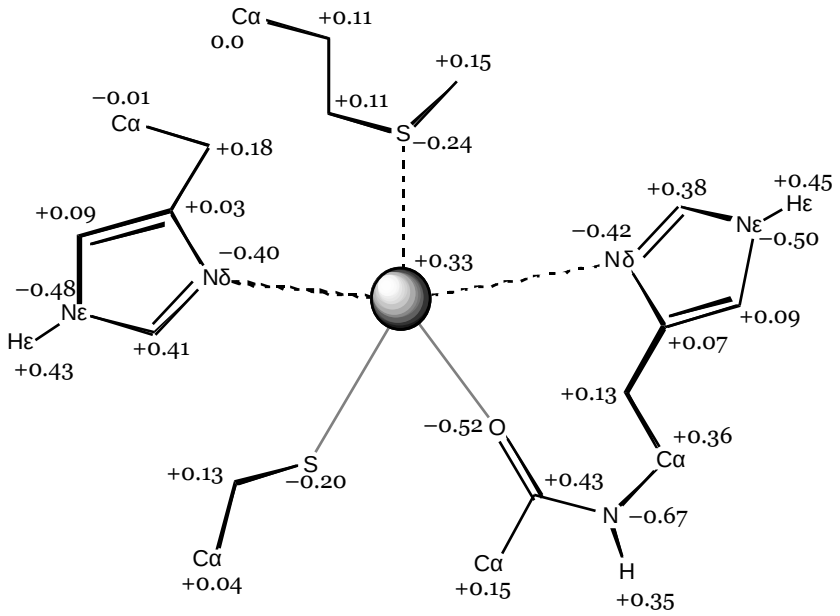


FIGURE 6.1.4. CHARGE DISTRIBUTION IN ACTIVE SITE (OXIDIZED STATE)

The inclusion of the protein charges does not have such a marked effect on the charge distribution. Although there are some subtle changes of a few hundreds a.u., globally the picture remains the same as in the *vacuo* case, although the standard deviations are a bit higher now. However, the redox difference, i.e. the difference between the reduced and oxidized state, remains virtually the same for all residues.

Discussion

The completely static inclusion of the protein environment can not account for the experimentally observed redox potential differences upon pH changes or site-directed mutations, as the protein is not allowed to relax in this section. For instance, the change in redox state is likely to induce changes in the protein-site interactions, e.g. through reorientation of polar groups. The solvent, which may also have a pronounced effect, has also not been taking into account in this section. However, it can be established if the presence of the protein environment has an effect on the active site energy or the charge distribution in it, or on both.

Energies

A direct connection between the total protein charge and the stabilization by the protein charges can be made. The ionization potentials for *wt2* azurin and the two subunits of F114A azurin with a protein charge of −3 are similar, as are the values for *wt1* azurin and the two subunits of F114A azurin with a protein charge of −2. At first sight, the values for N47D look odd, as they seem to indicate a total protein charge of −3, while in fact a value of −2 is found. But the N47D mutant is in fact a double mutant, with a second mutation at a position far away from the active site (Glu2→Gln). Therefore, the protein charge of the close by residues is indeed −3, as anticipated from the ionization potential change, while the total protein charge is the sum of this close by protein charge and the charge of residue 2.

The inclusion of the point charges leads to a stabilization of both the reduced and oxidized state, where the oxidized state is stabilized more than the reduced state (see Figure 6.1.5). Taking the reduced state *embedded* in the protein charges as reference state (*embed.red* in Figure 6.1.5), the additional stabilization when an electron is being removed in going to the *embedded* oxidized state (along path **C**), can be modeled quite accurately (average absolute deviation 0.7 kcal/mol along path **C**) by classical electrostatic interactions between the active site and the protein charges, when the MDC-q charges are used for the active site atoms.

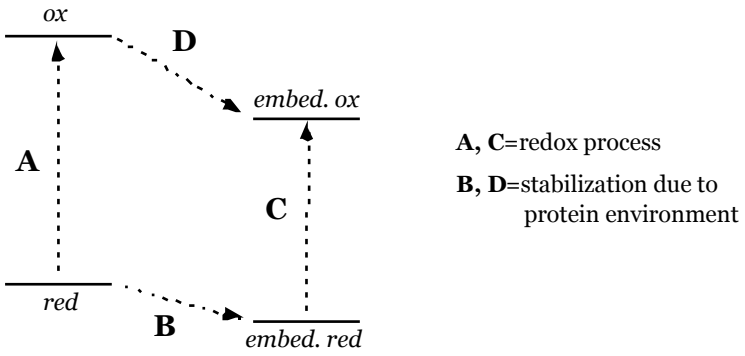


FIGURE 6.1.5. REDOX ENERGY PROFILE FOR EMBEDDING IN PROTEIN ENVIRONMENT

The stabilization of the reduced state due to the presence of the protein environment (path **B**) however can not be modeled by electrostatic interactions between the active site and protein charges alone; a subunit dependent *extra* stabilization of some 6-21 kcal/mol remains. This *extra* stabilization is most likely the result of polarization effects, which have not been taken into account in the classical electrostatics modeling.

The inclusion of protein charges therefore leads consequently to a lower energy, but the amount of stabilization depends on the azurin variant as well as on the (reduced or oxidized) state of the active site. This is just an energetic effect, which can be modeled quite well by classical electrostatics; the latter enables the use of classical MD simulations for obtaining redox potentials of copper proteins, when appropriate (MDC-q) charges for the active site are used.

Charge distributions

The r.m.s. variation of the (MDC-q) residue charges over the four azurin variants is rather small, especially if the differences in the active site geometries are taken into account. The distance of Gly45 to copper for instance varies from 2.3 to 3.1 Å; for Met121 it is in the range of 2.8-3.5 Å. Not only these weakly bonded axial groups show a considerable flexibility, also the distances of the in-plane ligands to copper vary to some extent. For His46, it is in the range from 2.0 to 2.2 Å, for Cys112 from 2.2-2.3 Å and for His117 from 1.9 to 2.6 Å. Still, the deviations in the charge distributions are rather small at 0.02-0.04 a.u.

The influence of including the protein charges in the DFT calculation is small; the changes in residue charges are at most some 0.1 a.u. However, again this change in residue charge is almost constant for all four azurin variants. The standard deviation increases just by a small amount. There is also hardly any difference between the two sets of protein charges; the residue charges after inclusion of either the Gromacs or MDC charges are rather similar. The difference between the reduced and oxidized state charges (redox difference) is

even more similar for the three environments (*in vacuo*, *Gromacs* or *MDC* protein charges). The difference between the three environments is at maximum some 0.02 a.u. This is a very promising result, as it means that in the MD simulations, the standard Gromacs charges can be used for the protein, while the MDC-q charges should be used for the active site.

The inclusion of protein charges does not have a marked effect on the charge distribution in the active site. There are some small differences, but they are the same for the four azurin variants.

Conclusions

The active site of the copper protein azurin has been studied with Density Functional Theory calculations for four variants: *wildtype* (*wt1* and *wt2*) and two mutants (N47D and F114A). The energy and charge distribution was computed along the process of removal of an electron from the system in steps of 0.2 electron. These calculations were performed either *in vacuo* or surrounded by the charges of the rest of the protein, for which two sets of charges were used, the standard Gromacs charges or charges obtained for amino acid residues from the Multipole Derived Charge analysis.

The ionization potentials of the active sites *in vacuo* are rather similar, with a deviation of at most 3 kcal/mol for the mutants. The ionization potential for *wildtype* azurin has been obtained for active site structures that have been determined experimentally at low (5.5) and high pH (9.0). This pH difference was found experimentally to cause a deprotonation of the His35 residue, as well as to have a negligible effect on the active site. It is confirmed by the DFT results, which give an ionization potential of 117.6 kcal/mol in both cases for the active site *in vacuo*. Because of the classical electrostatic effect, differences are observed when protein charges are included in the quantum calculation.

The inclusion of protein charges has a stabilizing effect on the active site, the energy is lowered by a subunit and state dependent amount. These stabilizations can be modeled quite accurately by classical electrostatics. The protein charges have a negligible effect on the charge distribution in the active site. This feature can be used in MD simulations, where the standard Gromacs charges can be used for the protein charges, while the MDC-q should be used for the atoms in the active site.

6.2 *Force constants*

Force field parameters for copper for several azurin molecules

In Section 6.1, Density Functional Theory[1] (DFT) was used to find the energy and charge distribution for the active site of a few azurin molecules. The resulting charges can be used directly in classical Molecular Dynamics (MD) simulations[160] for a proper description of the non-bonded interactions (together with standard Lennard-Jones parameters[127]). However, for the bonded interactions there are no standard force field parameters available, since the interactions of metals are difficult to generalize in terms of a simplified force field. Therefore, the force field parameters for describing the bonded interactions of copper with its surrounding residues should be obtained from DFT calculations.

As a first attempt, a simple approach has been tried, in which the DFT Hessian of the Cu atom alone (a *3x3* matrix with six independent values) was used to fit force constant values of copper-residue bonds. Five bonds connecting copper to the residues in the active site have been used, using a harmonic bond potential for each of them. The second derivative of these potentials with respect to the atomic coordinates provides a (force field, FF) Hessian matrix, which is a function of the force constants and the atomic coordinates (see Appendix of Section 3.2 for explicit formulas). The force constants for the five bonds were fitted to minimize the difference between the DFT and FF Cu-Hessians[a]. Although this procedure seems to give a reasonable Cu-Hessian for one *wildtype* structure, the obtained force constants look odd with a force constant for the "weakly bonded" Gly45 (at 2.96 Å) of 44 kcal/mol/Å^2, while for the in-plane ligand His46 (at 2.06 Å) a value of only 31 kcal/mol/Å^2 is found. Using this procedure for other active site structures even results in negative (!) force constants for some of the residues. Therefore, fitting the Cu-Hessian is not providing reliable force constants. Subsequently, a new method (IntraFF) has been developed, described in more detail in Section 3.2, which does take the bonding interactions into account on *both* sides, and provides reliable force constants. In the IntraFF method, only those parts of the DFT Hessian are needed that are related to the atoms involved in the bonds; as we want to have force constants for bonds of copper to the nearest atom in the residues (O-Gly45, N-His46, S-Cys112, N-His117, S-Met121), only the parts of the DFT Hessian related to these six atoms are needed (therefore a *18x18* matrix).[b]

Computational details

The Density Functional Theory calculations were performed with the ADF program (versions 2.3.3, 1999.03, 2000.02)[177,178,187] on a cluster of either IBM RS/6000 or Pentium-Linux boxes. In all calculations the Becke[120]-Perdew[121] exchange-correlation potential was used in a triple

[a] In principle, the DFT Hessian should be corrected for non-bonding (electrostatic and Lennard-Jones) interactions between copper and the residue atoms with which copper does not have bonding interactions; however, these corrections are of the order of 10^{-4} a.u. for the Hessian, while the DFT Hessian values for copper are of the order of 10^{-1} a.u.

[b] Of course one needs the complete active site residues in the DFT calculations; therefore, although the Hessian is obtained only for the Cu-O(Gly45)-N(His46)-S(Cys112)-N(His117)-S(Met121) atoms, it is obtained for a system in which all active site residue atoms are present

zeta basis set with polarization functions (TZP). As the Hessian is calculated in the ADF program[117,187] through numerical differentiation of the analytical gradients, obtaining the Hessian matrix takes a long time. Therefore, for each azurin variant one subunit was selected for which the Hessian was computed. Afterwards, the IntraFF analysis was performed using a standalone program.

In this section, the molecules used in Section 6.1 are considered, and some other azurin variants have been added. The additional molecules are the M121Q (Met121→Gln)[42] and N47L[84] (Asn47→Leu) mutants. The coordinates for the azurin variants were taken from PDB-files that were obtained from the Brookhaven Protein Database {F114A *Pseudomonas aeruginosa* (1azn), N47D *Pseudomonas aeruginosa* (1azr), wildtype *Pseudomonas aeruginosa* pH5.5 (4azu; hereafter referred to as *wt1*), wildtype *Pseudomonas aeruginosa* pH9.0 (5azu; *wt2*)} or from personal communication[280] {*Alcaligenes denitrificans* M121Q *Ad*, N47L *Ad*}.

Results

The IntraFF harmonic force constants for the azurin variants in both the reduced and oxidized state, with corresponding "equilibrium" bond distances are given in Table 6.2.1. The "equilibrium" bond distances were simply taken as the values as they are obtained in the X-ray data (which were obtained in the oxidized state).

TABLE 6.2.1. INTRAFF FORCE CONSTANTS *K* (KCAL/MOL/Å^2) AND DISTANCES *R* (Å)

	Gly45	*His46*	*Cys112*	*His117*	*Met/Gln-121*
wt1					
R(Cu-L)	2.955	2.064	2.267	1.978	3.164
K(red.)	26.3	67.2	148.9	124.1	30.6
K(ox.)	30.5	115.9	168.9	177.7	36.9
wt2					
R(Cu-L)	2.828	1.976	2.284	2.056	3.147
K(red.)	24.0	128.4	146.7	68.8	31.9
K(ox.)	28.0	181.7	166.9	117.8	39.3
n47d					
R(Cu-L)	2.362	2.087	2.218	1.869	3.440
K(red.)	21.2	61.3	199.6	246.6	20.8
K(ox.)	32.9	109.4	223.5	304.6	33.3
f114a					
R(Cu-L)	3.066	2.208	2.216	2.362	3.007
K(red.)	14.1	72.5	179.3	16.2	36.1
K(ox.)	16.5	89.3	179.0	17.5	39.1
n47l					
R(Cu-L)	2.951	1.975	2.130	1.943	3.121
K(red.)	39.9	164.5	290.4	169.1	32.7
K(ox.)	49.2	226.5	323.8	229.5	45.3
m121q					
R(Cu-L)	3.468	1.907	2.134	2.064	2.250
K(red.)	40.3	243.1	298.2	82.7	37.7
K(ox.)	59.2	300.8	324.3	124.7	55.3

Discussion

The IntraFF force constants for the copper to residue bonds are consistently larger in the oxidized state than in the reduced state, indicating stronger bonding in the former. Furthermore, the force constants for the in-plane ligands are always larger than for the axial groups, except for His117 in the F114A mutant. However, in that case the Cu-ligands is elongated to a large extent at 2.36 Å.

The Cu-S(Cys112) force constant of ~170 kcal/mol/Å^2 as well as the Cu-N force constants of 115-180 kcal/mol/Å^2 for *wildtype* azurin are small compared to "normal" bonds. For instance, the force constant for a C-C single bond is 620 kcal/mol/Å^2 and for a double bond even 938 kcal/mol/Å^2 in the AMBER95[126] force field, while a C-S bond has a force constant of 474 kcal/mol/Å^2. Still, the Cu-Cys/His force constants are considerably larger than the Cu-Gly45/Met121 bonds (30-40 kcal/mol/Å^2). The latter values are still too large to be represented by non-bonded interactions only in force field calculations; non-bonded interactions between copper and the active site residues result in Hessian values of $1–5\cdot10^{-4}$ a.u. while force constant values of ~30 kcal/mol result in Hessian values $1\cdot10^{-2}$ a.u.

For the in-plane ligands (His46, Cys112 and His117) an interesting trend can be observed: the force constant increases if the "equilibrium" distance decreases, which is indicative for an anharmonic potential (for a pure harmonic potential the force constant would remain constant). For instance, in Figure 6.2.1 the force constants for His117 are plotted against the "equilibrium" distance.

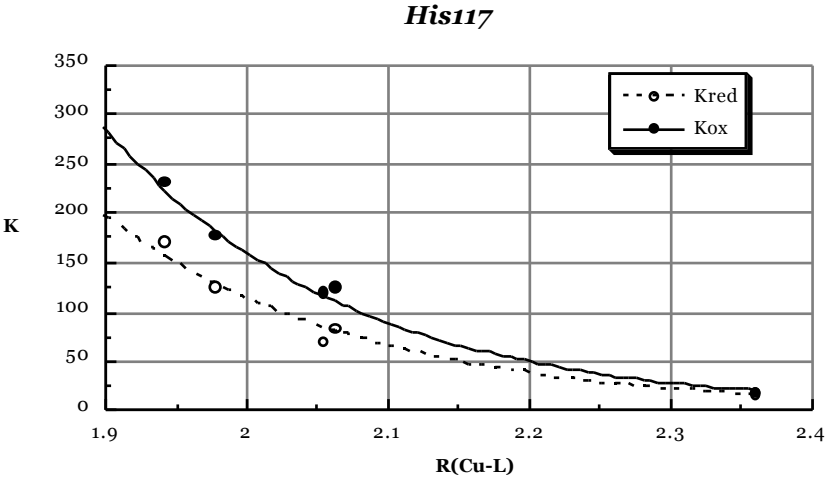


FIGURE 6.2.1. HIS117 INTRAFF FORCE CONSTANTS AS A FUNCTION OF THE DISTANCE

Fitted through the points on the plot is an exponential function, which fits nicely with the obtained points. A similar exponential function would have been found for the Hessian (in 1D) of either the Morse[130] or Frost[131,132] potential (see Section 3.2), which are both anharmonic bond potentials. Using the IntraFF method, Frost parameters were obtained for the five bonds in the four azurin variants; the average values of these parameters are given in Table 6.2.2 for both the reduced and oxidized state, together with the equilibrium distance and dissociation energy. Also given is the force constant for a harmonic potential at the Frost equilibrium distance, which is equal (in 1D) to the 2nd derivative of the Frost potential at the equilibrium distance.

TABLE 6.2.2. FROST PARAMETERS (*Z1*,*Z2*,*a*,*b*; A.U.) FOR ACTIVE SITE OF AZURIN *IN VACUO*, WITH CORRESPONDING EQUILIBRIUM DISTANCE R_{EQ} (Å), DISSOCIATION ENERGY D_E (EV) AND FORCE CONSTANT $K(R_{EQ})$ FOR HARMONIC POTENTIAL (KCAL/MOL/Å^2)

	Z1	*Z2*	*a*	*b*	R_{eq}	D_e	$K(R_{eq})$
Gly45							
reduced	29	8	1.37	44.8	3.08	0.05	9.1
oxidized	29	8	1.17	51.4	2.78	0.42	62.4
His46							
reduced	29	7	1.60	65.3	1.93	0.76	214.9
oxidized	29	7	1.51	65.1	1.95	1.03	263.4
Cys112							
reduced	29	16	1.51	130.2	2.19	0.96	237.3
oxidized	29	16	1.44	133.7	2.15	1.50	346.1
His117							
reduced	29	7	1.62	65.3	1.93	0.70	203.7
oxidized	29	7	1.53	65.6	1.93	1.03	269.0
Met121							
reduced	29	16	1.10	94.0	3.03	0.63	83.9
oxidized	29	16	1.09	92.4	3.08	0.61	78.7

Conclusions

Force field parameters for the bonding interactions of copper in a few azurin variants have been obtained from Density Functional Theory calculations. The force constants were obtained with the IntraFF method, which extracts them from a quantum chemically computed Hessian matrix. This method provides reliable force constants that can be used in classical Molecular Dynamics simulations.

The obtained force constants are consistently larger in the oxidized state than in the reduced state, indicating stronger bonding in the former. Furthermore, the force constants show anharmonic behavior, i.e. the force constant values increase with decreasing “equilibrium” distance. The parameters for the anharmonic Frost potential have been obtained with the IntraFF method for all azurin variants and averaged to obtain a generally applicable copper (azurin) force field.

6.3 *Axial bonding*

Checking the bonding character of the axial groups and monitoring the potential energy surface

In this section, the bonding character of the axial groups is investigated by looking at the mixing of the copper and axial residue orbitals. The residue-copper distance of either the Met121 and the Gly45 residue has been systematically varied from 2.5 to 4.0 Å in steps of 0.1 Å, while the rest of the geometry was kept fixed. This enables a check on the influence of this distance on the mixing of copper-residue orbitals, and a potential energy surface is obtained for these two ligands, both in the reduced and the oxidized state.

Computational details

The active site structure[a] used in this study was taken from the PDB-file of *wildtype Pa* azurin (B subunit) for which the X-ray data were obtained at pH 5.5[7]. The copper-residue distances for the residues in the active site are 2.064 Å (His46), 2.267 (Cys112), 1.978 Å (His117) for the in-plane ligands, and 2.955 Å (Gly45), 3.164 Å (Met121) for the axial ligands. The coordinates were transformed from Cartesian format into Z-matrix format[1]. This enables to vary either the Cu-O(Gly45) or the Cu-S(Met121) distance, and let the corresponding residue follow. In the following, either the Cu-Gly45 or the Cu-Met121 distance was varied, while the rest of the coordinates were kept fixed. As the Gly45 and His46 residues are connected, varying the Cu-Gly45 could have resulted in changing the Cu-His46 distance also (see Figure 6.3.1).

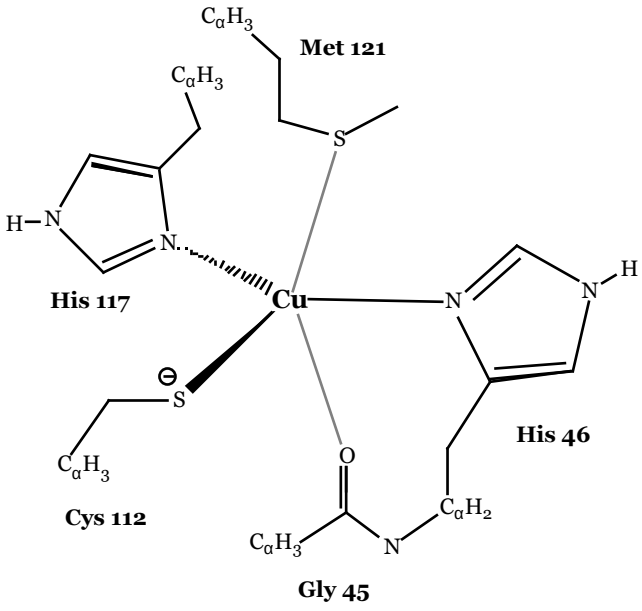


FIGURE 6.3.1. ACTIVE SITE OF AZURIN

[a] The active site consists of the same residues and atoms as described in section 6.1

In the used Z-matrix however, only the atoms of Gly45 changed position, while all atoms belonging to His46 remained at their position. The change in Cu-Gly45 distance was therefore accompanied by a change in the angles between the backbone of His46 and the atoms of Gly45.

The Density Functional Theory calculations were performed with the ADF program (version 2000.02)[177,178,187] on a cluster of Pentium-Linux boxes. In these calculations the Becke[120]-Perdew[121] exchange-correlation potential was used in a triple zeta basis set with polarization functions (TZP).

Results and discussion

Variation of the Gly45-Cu distance

For the reduced state with Gly45 at 2.5 Å, the highest occupied molecular orbital (HOMO), which is doubly occupied, consists for 51 % of contributions from the sulphur of Cys112 and 36 % from copper, and no contribution from the Gly45 oxygen. This oxygen has some contributions in orbitals where it is mixed with orbitals of its neighboring atoms, and a very small mixing (1 %) with Cu d-orbitals in one orbital (HOMO-2[a]). When Gly45 is placed at the other extreme of 4.0 Å, the HOMO has hardly changed: 54 % S_{112} and 33 % Cu. Naturally, no mixing with copper orbitals are found at this distance.

For the oxidized state with Gly45 at 2.5 Å, the singly occupied HOMO consists again mainly of contributions from S_{112} (52 %) and copper (28 %), and no contribution from O_{45}. The oxygen atom however does have some small contributions in the HOMO-1 (1 %), and some more significant contributions in the HOMO-3 (14 %) that consists for 53 % of Cu d-orbitals, and the HOMO-10: 50 % O_{45} and 18 % Cu d-orbitals. With the Gly45 at 4.0 Å, the HOMO consists mainly of copper d-orbitals (26 %) and S_{112} orbitals (52 %). At this distance, the oxygen of Gly45 does of course not mix with the copper orbitals anymore.

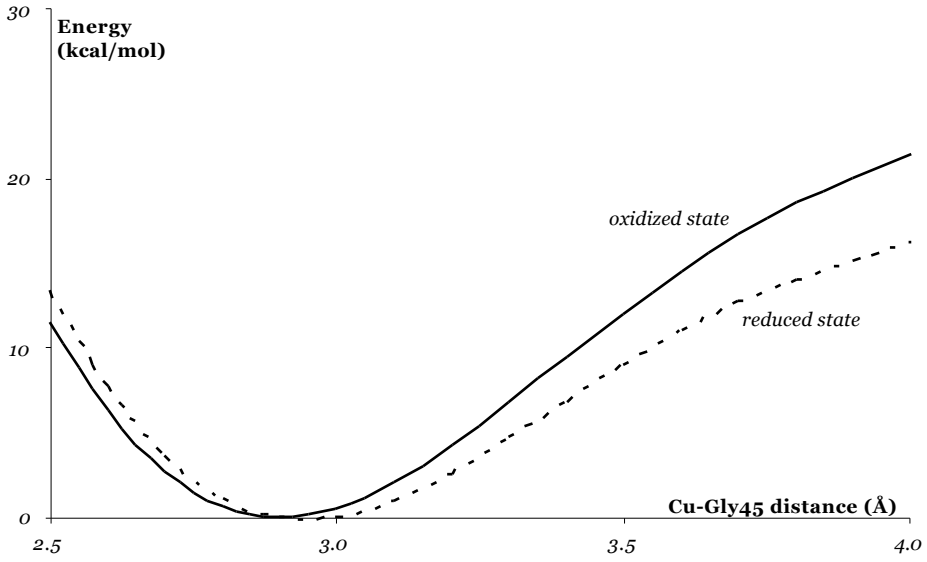


FIGURE 6.3.2. ENERGY PROFILE FOR GLY45-CU DISTANCE

[a] The HOMO-2 means the second orbital below the Highest Occupied Molecular Orbital (HOMO)

The energies for the reduced and oxidized state as function of the Gly45-Cu distance are given in Figure 6.3.2. The energy is given for the reduced and oxidized state relative to the lowest energy of that state, which is found in both cases at a Cu-Gly45 distance of 2.9 Å. The energy needed to put the Gly45 residue at 2.5 Å is 13.4 and 11.6 kcal/mol for the reduced and oxidized state respectively, while it costs 16.3 (*reduced state*) and 21.6 (*oxidized state*) kcal/mol to increase the Cu-Gly45 distance from 2.9 to 4.0 Å.

Variation of the Met121-Cu distance

For the reduced state with Met121 at 2.5 Å, the doubly occupied HOMO of the reduced state consists of contributions from S_{112} (56 %) and copper (31 %). The sulphur of Met121 is not involved in the HOMO, but does have some contribution to the HOMO-1 (5 %), which consists further of contributions from copper (41 %) and S_{112} (37 %). As expected, no such mixing in of S_{121}-orbitals is observed when Met121 is placed at 4.0 Å from copper; only contributions to orbitals located entirely on Met121 are found.

In the oxidized state with Met121 at 2.5 Å, the singly occupied orbital consists of contributions from Cys112 (37 %), copper (31 %) and Met121 (16 %). Also in other molecular orbitals considerable contributions from Met121 are found. With Met121 at 4.0 Å, the singly occupied HOMO consists mainly of contributions of S_{112} (51 %) and copper (27 %) with some smaller contributions (~2 %) from the neigboring nitrogens of His46 and His117.

The energies for the reduced and oxidized state as function of the Met121-Cu distance are given in Figure 6.3.3. The energy is given for the reduced and oxidized state relative to the lowest energy of that state, which is found at a Cu-Met121 distance of 3.6 Å for the reduced state and at 3.5 Å for the oxidized state. The energy needed to put Met121 at 2.5 Å is much larger than for Gly45: 22.3 kcal/mol for the reduced state and 29.8 kcal/mol for the oxidized state.

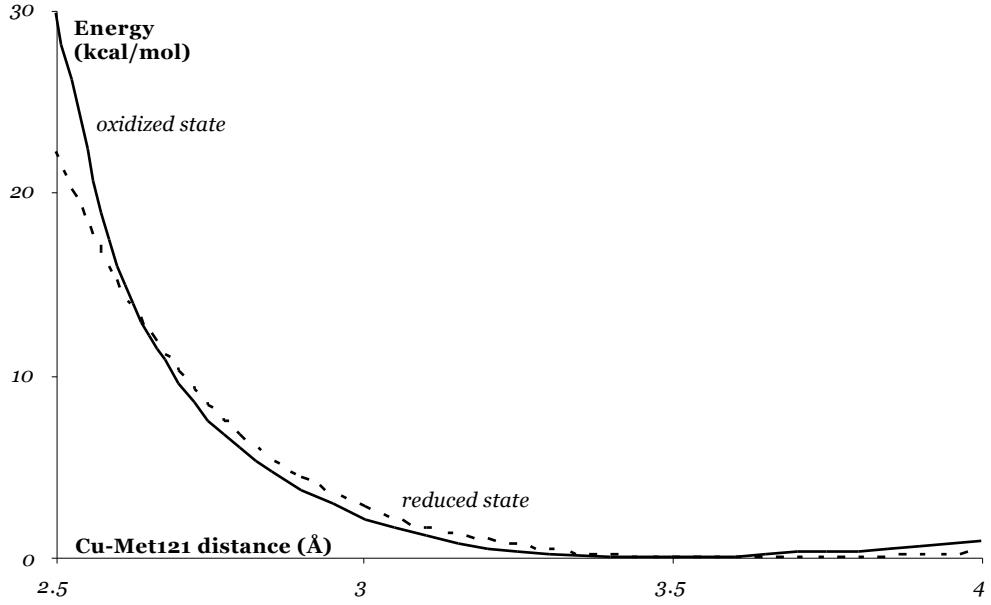


FIGURE 6.3.3. ENERGY PROFILE FOR MET121-CU DISTANCE

Around the lowest energy distance, a rather weak bonding profile is found; for instance in the oxidized state, it requires only 1.2 kcal/mol to decrease the Cu-Met121 distance from 3.5 to

3.1 Å, and only 0.9 kcal/mol to increase it to 4.0 Å. Therefore, a plateau is observed where the Met121 can move for almost an Angstrom with an energy change of approximately only 1.0 kcal/mol.

The observed plateau makes it difficult to get an accurate geometry for the active site by treating it *in vacuo*, as the influence of the protein on the Met121 residue is probably of the same order of magnitude as the energy difference at this plateau, or even higher. The inclusion of the protein in QM/MM calculations (see Section 2.3 and Chapter 9) is therefore needed to obtain correct geometries.

Lowest energy orbital compositions

In the reduced state, both with Gly45 placed at 2.9 Å from copper and with Met121 placed at 3.6 Å, the doubly occupied HOMO consists for 51 % of S_{112} contributions and 36 % of copper contributions. At these distances, no significant mixing between copper and oxygen orbitals is found, while some small mixing is observed between Met121 and copper orbitals.

In the oxidized state for the singly occupied HOMO, the contributions of S_{112} and copper are respectively 52 % and 27 %. Some small mixing is observed between both copper and oxygen orbitals, and copper and sulphur (Met121) orbitals.

Force field parameterization

Mixing of copper and residue orbitals is indicative of covalent interactions between copper and the residues. At equilibrium distance, no mixing of Gly45 and copper orbitals is observed in the reduced state, while there is some mixing in the oxidized state; Met121 and copper orbitals mix both in the reduced and oxidized state. Therefore, it seems there are covalent interactions between copper and Met121 in both states, between copper and Gly45 only in the oxidized state. This is reflected in the "harmonic Frost" force constants (Section 6.2), which show the same pattern; therefore, modeling the interactions in the active site of azurin with five copper-residue bonds seems justified.

Conclusions

The influence of the distance of the axial groups from the copper is investigated by looking at the potential energy surface obtained by monitoring the energy as each one of them is moved away from the copper in steps of 0.1 Å. At the same time, the amount of mixing of the orbitals of the copper and the axial ligands is monitored.

The energy surface for the Gly45 residue shows a well defined minimum around 2.9 Å, both for the reduced and oxidized state, which is very close to where it is found in the crystal structures of azurin. On the contrary for the Met121 residue a very flat energy surface is found, with a change in energy of around 1 kcal/mol if the residue is moved from 3.0 to 4.0 Å. The "optimal" copper-Met121 distance is then for both the reduced and oxidized state around 3.5 Å. This enhanced freedom to move at almost no cost may presumably be the result of the absence of the protein matrix.

The amount of mixing between copper orbitals and orbitals of the oxygen of Gly45 is small, although some mixing is present when the latter is at a distance of 2.5 Å from copper. Some mixing is observed between the copper orbitals and the orbitals of the sulphur of Met121, both in the reduced and oxidized state.

6.4 *Vibrational frequencies*

Validating the copper force field[a]

In Section 6.2, the computed Hessians from Density Functional Theory calculations[1] were used to extract force constants for bonds of copper to the residues in the active site. Those Hessians provide even more information; they can be used to obtain vibrational frequencies with corresponding intensities.[1] For this, the Hessian H has to be transformed into mass-weighted coordinates:

$$H_{ij}^{mass\text{-}weighted} = \frac{H_{ij}}{\sqrt{M_i M_j}} \qquad (1)$$

with M_i the mass of atom i. After diagonalization of the mass-weighted Hessian, the eigenvalues λ_i can be transformed quite easily into the frequencies ν_i with corresponding normal mode vectors:

$$\nu_i = \frac{1}{2\pi}\sqrt{\lambda_i} \qquad (2)$$

This procedure works for any Hessian in Cartesian coordinates. Therefore, it can be used directly to check the vibrational frequencies that would be the result of the interactions of the five bonds (see Section 6.2).

The vibrational frequencies obtained in that way would refer to an isolated system containing only those five bonds. However, they will be applied to the active site in a protein and one wants to compare them with experimentally observed frequencies, which are usually obtained at finite (room) temperature. For a meaningful comparison, one therefore should perform Molecular Dynamics simulations[160] at that temperature and extract the values for the frequencies from those simulations. In principle, the same procedure could be used as described above, i.e. calculate the Hessian for all coordinates, transform it to the mass-weighted coordinate system and diagonalize it. However, this would be an enormous task; the Hessian matrix should be averaged over time, while the time needed for just one Hessian would be enormous due to the system size of several thousands of atoms. A more efficient way of extracting the values for the vibrational frequencies can be obtained by monitoring the copper-residue bond lengths in time. Transforming the values from real space to frequency space by taking the Fourier transform results then directly in the values for the vibrational frequencies resulting from those bonds. As these values are obtained in simulations at a particular temperature, they can be compared directly with the experimentally observed values.

---

[a] Joint study by M. van den Bosch (MD simulations) and M. Swart (DFT calculations)

Computational details

The Density Functional Theory calculations were performed with the ADF program (versions 2.3.3, 1999.03, 2000.02)[177,178,187] on a cluster of either IBM RS/6000 or Pentium-Linux boxes. In all calculations the Becke[120]-Perdew[121] exchange-correlation potential was used in a triple zeta basis set with polarization functions (TZP). The Hessian matrix (a matrix containing the second derivatives of the energy with respect to atomic coordinates) was determined for copper and its neighboring atoms in the five residues in the active site. In total the Hessian matrix has therefore a dimension of *18x18*, and leads to 18 vibrational frequencies. For a molecule consisting of 6 atoms (and therefore also to 18 vibrations), six of these would correspond to translation or rotation of the molecule and have a frequency of zero; however, in this case we are dealing with a sub-part of the Hessian of the molecule (which has a size of *183x183* for the active site consisting of 61 atoms). I.e., the system of the six atoms is connected to the residues and therefore does not have the six modes with frequency zero.

The coordinates for the azurin variants were taken from PDB-files that were obtained from the Brookhaven Protein Database {F114A *Pseudomonas aeruginosa* (1azn), N47D *Pseudomonas aeruginosa* (1azr), wildtype *Pseudomonas aeruginosa* pH5.5 (4azu; hereafter referred to as *wt1*), wildtype *Pseudomonas aeruginosa* pH9.0 (5azu; *wt2*)} or from personal communication[280] { *Alcaligenes denitrificans* M121Q AD[42], N47L[84] AD}.

The Molecular Dynamics simulations and analyses were performed with the Gromacs suite[279] (version 2.1) on a SGI R12000 and a DEC/Alpha-linux cluster. The Multipole Derived Charges[183] were used for the active site atoms and the IntraFF force constants for *wildtype Pa* azurin (see Section 6.2) for bonds of copper to the five residues in the active site. For all other force field parameters, the standard GROMOS96[127] values were used, where the high Lennard-Jones[160] repulsion parameter for Cu[127] was used throughout. The simulations consisted of 1 nanosecond at a temperature of 300K and a pressure of 1 bar, which were kept constant using the Berendsen[166] thermo-barostat. The protein was solvated in 4152 (SPC/E[281]) water molecules. The cut-off radius was set at 14 Å, used for Lennard-Jones and electrostatic interactions, and 8 Å for the neighborlist that is updated every ten time steps. No constraints were put on bonds, angles or dihedrals in the active site, while the bond lengths were constrained for the rest of the protein, and the bond lengths and angle for the solvent molecules using the LINCS procedure.

Results

The vibrational frequencies from the DFT Hessians are given in Table 6.4.1 for both the reduced and oxidized state. Given also are the intensities for the frequencies, which are obtained from taking the derivative of the dipole moment with respect to the atomic coordinates. The frequencies for the reduced states of the six azurin variants are mostly lower than those found for the oxidized state. This is in agreement with the force constants for the copper-residue bonds, which were extracted from these Hessians and found to be smaller in the reduced than in the oxidized state. For all but the F114A-mutant, an intense frequency around 400 cm^{-1} is observed (intensity around 30 km/mol), while even more intense frequencies are observed in the 900-1100 cm^{-1} region. Contributions of the Cu-Cys

bond to the normal mode vectors[a] are mainly found in the 280-370 cm^{-1} region, with some smaller contributions for the frequencies in the region of 400 to 500 cm^{-1}.

TABLE 6.4.1. FREQUENCIES (CM^{-1}) AND *INTENSITIES* (KM/MOL) FROM DFT

wt1		*wt2*		*n47d*		*f114a*		*n47l*		*m121q*	
					reduced state						
84	*1.8*	86	*2.9*	67	*2.3*	56	*4.4*	82	*2.0*	93	*2.6*
*109	*9.7*	*117	*6.9*	*89	*9.8*	70	*3.8*	84	*8.5*	103	*7.6*
*124	*1.2*	*128	*3.5*	*119	*1.4*	*119	*1.3*	*157	*0.5*	*126	*0.9*
*158	*0.3*	*157	*0.6*	190	*1.3*	142	*11.1*	169	*1.2*	*189	*1.7*
196	*1.3*	180	*1.3*	*195	*0.3*	189	*1.0*	*173	*0.7*	211	*6.9*
218	*7.2*	223	*7.4*	229	*6.4*	293	*7.6*	238	*5.6*	*240	*6.5*
300	*12.4*	**292**	*13.2*	**313**	*8.7*	**326**	*19.4*	**369**	*6.7*	**380**	*4.7*
406	*4.2*	417	*33.0*	412	*30.5*	383	*1.4*	409	*33.4*	*400	*27.2*
417	*25.0*	*425	*2.3*	448	*3.5*	*390	*11.8*	420	*3.5*	*405	*2.4*
425	*9.6*	*436	*14.8*	**450**	*13.3*	411	*1.8*	**452**	*12.7*	417	*27.0*
*435	*14.3*	460	*0.2*	454	*0.2*	414	*5.3*	462	*0.5*	435	*8.8*
453	*0.2*	471	*3.9*	481	*0.2*	417	*2.3*	491	*0.5*	**445**	*10.6*
476	*0.3*	491	*0.4*	507	*4.8*	426	*20.5*	529	*3.9*	1071	*10.2*
1056	*5.6*	691	*3.9*	929	*3.1*	910	*65.3*	929	*1.1*	1084	*8.7*
1078	*12.7*	1067	*10.0*	954	*4.3*	914	*43.3*	1056	*113.6*	1086	*61.6*
1090	*19.3*	1075	*5.8*	1066	*18.1*	995	*18.4*	1078	*13.1*	1103	*99.5*
1090	*89.2*	1080	*102.2*	1070	*20.4*	1091	*46.8*	1088	*2.7*	1104	*5.9*
1097	*6.3*	1095	*7.3*	1090	*95.0*	1117	*17.7*	1101	*8.7*	1114	*6.0*
					oxidized state						
93	*3.2*	93	*5.8*	90	*6.9*	*57	*10.2*	97	*6.1*	100	*4.5*
118	*3.0*	123	*1.1*	104	*1.3*	92	*8.9*	108	*1.3*	119	*3.6*
*151	*11.4*	*154	*5.2*	*145	*12.8*	*129	*4.2*	179	*1.5*	*158	*5.9*
*183	*6.0*	*181	*4.9*	195	*0.7*	136	*1.4*	*186	*3.9*	*207	*5.8*
203	*0.8*	189	*2.1*	*212	*5.6*	191	*1.5*	195	*4.5*	*209	*3.4*
215	*7.5*	220	*6.3*	228	*5.6*	288	*7.8*	237	*5.6*	*232	*11.5*
296	*7.9*	**287**	*10.4*	**313**	*8.2*	**320**	*6.9*	**366**	*7.4*	**375**	*4.9*
*408	*0.1*	416	*30.6*	411	*39.1*	368	*1.4*	411	*32.4*	*406	*31.2*
*418	*31.2*	*418	*1.9*	**433**	*0.4*	387	*5.0*	424	*2.3*	*408	*6.2*
*420	*6.3*	431	*0.9*	448	*1.5*	409	*4.2*	**444**	*2.3*	417	*23.2*
428	*1.3*	458	*2.1*	454	*2.6*	414	*0.9*	460	*0.3*	**436**	*2.1*
451	*4.0*	485	*4.4*	480	*2.5*	416	*1.7*	490	*2.8*	442	*5.9*
476	*2.9*	516	*13.9*	512	*2.5*	424	*25.7*	531	*7.2*	1085	*14.1*
1080	*13.8*	711	*9.4*	945	*22.1*	912	*51.4*	958	*12.3*	1096	*53.1*
1090	*85.7*	1080	*56.1*	985	*4.9*	918	*72.9*	1059	*137.8*	1101	*97.8*
1093	*76.5*	1082	*79.6*	1076	*17.0*	1003	*22.2*	1093	*8.0*	1105	*119.1*
1106	*0.4*	1096	*25.3*	1088	*36.3*	1099	*39.7*	1108	*5.2*	1120	*7.7*
1109	*16.7*	1104	*2.6*	1099	*143.2*	1129	*21.4*	1113	*16.5*	1138	*14.0*

Bold frequencies have more than 40 % contributions of the Cu-Cys(112) bond[a]
Frequencies indicated with an asterisk (*) have contributions between 10 and 40 %

[a] Obtained by taking the overlap of the normal mode vector and the Cu-Cys112 bond vector

The frequencies as obtained from the FF Hessian matrices (coming from the five copper-residue bonds only; see Section 6.2) are given in Table 6.4.2; in this case, only frequencies can be obtained. The frequencies from the five classical bonds are observed in the low-frequency region (100-570 cm^{-1}), both for the reduced and the oxidized state. As the force constants have a higher value in the oxidized state, the frequencies are found at higher values than in the reduced state. The contribution of the Cu-Cys bond to the normal mode vectors is found mainly in the 270-420 cm^{-1} region.

TABLE 6.4.2. FREQUENCIES (CM^{-1}) FROM THE CLASSICAL COPPER-RESIDUE BONDS

wt1	*wt2*	*n47d*	*f114a*	*n47l*	*m121q*
		reduced state			
121	121	101	107	128	170
162	158	143	126	195	205
*246	**249**	*244	146	**355**	288
296	**291**	**332**	*258	422	**391**
*363	**371**	*508	**329**	**447**	**515**
		oxidized state			
132	132	128	114	149	206
175	172	179	131	217	249
287	**286**	**313**	152	**390**	*349
357	**356**	**366**	**279**	493	**415**
*432	*438	*564	**339**	**504**	**569**

Bold frequencies have more than 40 % contributions of the Cu-Cys(112) bond
Frequencies indicated with an asterisk (*) have contributions between 10 and 40 %

The Molecular Dynamics simulations were performed for *wt1* azurin only in either the reduced or oxidized state. In these simulations, both the active site and the protein are stable (see Figures 6.4.1 and 6.4.2).

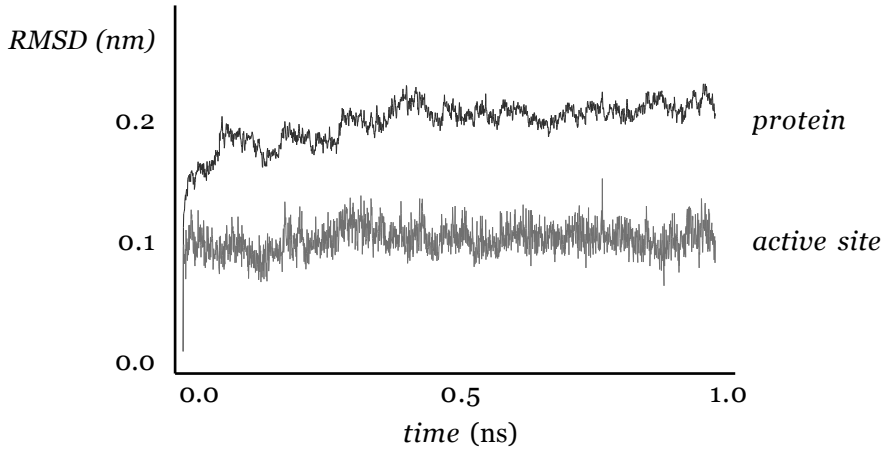


FIGURE 6.4.1. RMSD OF REDUCED STATE DURING MD SIMULATION

In these figures, the root mean square displacement (RMSD) of the atoms from their initial positions is given as function of the simulation time, either for the active site or the protein.

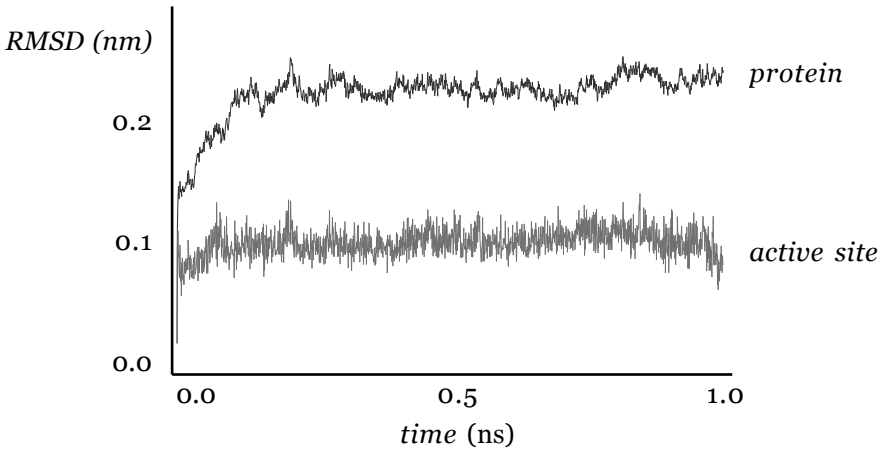


FIGURE 6.4.2. RMSD OF OXIDIZED STATE DURING MD SIMULATION

The RMSD of the active site atoms is in both the reduced and the oxidized state just below 1 Å, while it is around 2 Å for the C-alpha atoms of the complete protein. The stability of the site is confirmed by the average copper-residue distances as obtained in the MD simulation over a period of 500 ps (see Table 6.4.3). The difference between the experimental and simulated distances is small (less than 0.1 Å) and at most of the same magnitude as the experimental uncertainty.

TABLE 6.4.3. AVERAGE COPPER-RESIDUE DISTANCES (Å) FROM MD SIMULATIONS

	reduced state		*oxidized state*	
	X-ray	*MD*	*X-ray*	*MD*
R(Cu-Gly45)	*3.02 ± 0.07*	3.04	*2.97 ± 0.09*	3.05
R(Cu-His46)	*2.13 ± 0.09*	2.09	*2.08 ± 0.06*	2.08
R(Cu-Cys112)	*2.29 ± 0.01*	2.31	*2.24 ± 0.04*	2.29
R(Cu-His117)	*2.09 ± 0.09*	2.07	*2.01 ± 0.07*	2.07
R(Cu-Met121)	*3.25 ± 0.07*	3.34	*3.15 ± 0.07*	3.29
R(Cu-N_2S plane)	*0.09*	0.12	*0.08*	0.08

The vibrational frequencies for the Cu-Cys and both Cu-His bonds as obtained in the MD simulations[a], are given in Figure 6.4.3. The vibrational frequencies are all found below the 450 cm^{-1}, where the Cu-His frequencies are found mainly in the 200-350 cm^{-1} region and the Cu-Cys frequencies around 280 cm^{-1} and mainly from 350-400 cm^{-1}. They compare very well with the frequencies observed in Resonance Raman spectra of *wildtype* azurin (see Figure 6.4.3).

[a] from smoothed Fourier transform

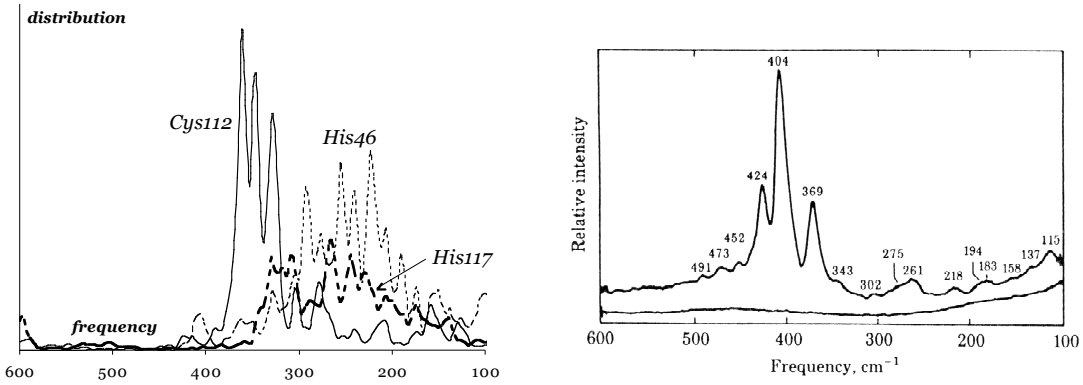


FIGURE 6.4.3. VIBRATIONAL FREQUENCIES FOR CU-RESIDUE BONDS FROM MD SIMULATIONS (LEFT) AND RESONANCE RAMAN SPECTRUM OF *WILDTYPE* AZURIN (TAKEN FROM REF 62)

Discussion

The frequencies obtained in Density Functional Theory[1] calculations are all found below 1200 cm^{-1} (see Table 6.4.1), with the most intense ones observed around 400 cm^{-1} and above 900 cm^{-1}. The normal mode vectors for the frequencies with higher values (> 900 cm^{-1}) consist exclusively of contributions from copper, nitrogen and oxygen atoms. Some preliminary tests where classical N-Cu-O angles have been added to the force field seem to indicate that they may be partly responsible for these higher frequencies.

The highest frequency where a sulphur atom is contributing to the normal mode vector is observed at 516 cm^{-1} (*wt2 in oxidized state*; Met121 sulphur). The largest contributions to a normal mode vector by the Cu-Cys bond are observed primarily in the region of 280-370 cm^{-1}, with some smaller contributions observed for frequencies between 400-500 cm^{-1}. Depending on the copper-sulphur distance, these frequencies shift up- or downwards. For instance, for *wt1* and *wt2* azurin with a Cu-S distance of ~2.27 Å, the seventh frequency is found around 290 cm^{-1}, while for the N47D and F114A mutants (with a smaller Cu-S distance of 2.22 Å) it is increased to around 320 cm^{-1} and for the N47L and M121Q mutants with an even smaller distance of 2.13 Å the frequency occurs around 370 cm^{-1}.

A similar but less pronounced trend is observed for the average of the Cu-His distances in a certain azurin variant and the thirteenth frequency. For the N47L mutant with an average Cu-N distance of 1.96 Å it is observed at 531 cm^{-1}, while for the F114A mutant with an average of 2.29 Å it is found at only 424 cm^{-1}. It is found inbetween these two extremes for *wt1*, *wt2* and the N47D mutant. For M121Q-azurin, it is found at 1085 cm^{-1} due to mixing in of the Cu-Gln121 bond, which has a distance of only 2.25 Å.

The frequencies from the five copper-residue bonds alone are all lower than 570 cm^{-1} (see Table 6.4.2). The positioning of these frequencies depends to a large extent on the Cu-residue distances, as can be seen from the frequencies observed in either the N47D or the F114A mutant. In the former, the in-plane copper-ligand distances are relatively small leading to higher values for the observed frequencies. The latter, with rather large copper-His distances, exhibits relatively low frequencies. In all cases however, the highest frequency has a considerable contribution from either one of the Cu-His bonds, with an usually smaller contribution from the Cu-Cys bond. The second highest observed frequency consists mainly of contributions from the Cu-Cys bond and Cu-His bonds, while the third frequency (around 290 cm^{-1}) consists mainly of contributions from the Cu-Cys bond. The two lowest

frequencies consist mainly of contributions from the axial residues, with some small contributions from either the Cu-Cys or the Cu-His bonds.

In the Molecular Dynamics simulations of 1 ns, both the active site and the protein remain stable after an initial equilibration period of some 500 ps (reduced state) or 200 ps (oxidized state). In both cases, the active site reaches stabilization faster then the protein, where a smaller root mean square deviation is observed for the oxidized state than for the reduced state. As the protein structure was determined experimentally for the former, this does not come as a surprise. The copper-residue distances observed during the second half of the 1 ns simulations are found close to the ones found in the X-ray data. For the in-plane ligands, a good agreement is found between simulated and experimental distances. For the axial residues the differences are somewhat larger, but remain within the experimental accuracy of 0.1-0.2 Å of the X-ray data.

The distance of the copper from the N_2S plane shows a good agreement between the simulated and experimental values. In both experiment and simulation, the copper remains above the plane at some 0.1 Å. As this distance has not been modeled in any kind of bonding interactions, it may serve as valuable check for determining the validity of the copper force field. The good agreement between experimental and simulated values indicates that the copper force field described in Sections 6.1 and 6.2 gives a proper description for the bonding interactions in the active site of azurin.

This is confirmed by the vibrational frequencies that are obtained in the MD simulations. For the Cu-Cys bond, they are observed mainly in the 300-400 cm^{-1} region with some smaller ones observed around 280 cm^{-1}, while for the Cu-His bonds they are observed at (mainly) lower or higher frequencies (see Figure 6.4.3). Although this figure doesn't provide intensities, as there is a large resemblance between the frequencies and normal modes observed in DFT calculations, the copper force field and the MD simulations, it seems justified to use the intensities obtained in the DFT calculations and apply them to the frequencies in a particular frequency range. Therefore, the intense frequencies observed in Resonance Raman studies around 400 cm^{-1}, which were attributed to copper-sulphur interactions, show up also in either DFT, copper force field or MD simulations results.

Conclusions

This section handles the vibrational frequencies that are obtained from either Density Functional Theory (DFT) calculations, copper force field calculations where the force field parameters have been extracted from Hessian matrices obtained in the DFT calculations, or from Molecular Dynamics simulations. A general agreement is observed for the frequencies resulting from the copper-residue bonds, which are all found below 1200 cm^{-1}.

In the DFT calculations, the most intense frequencies are observed around 400 cm^{-1} and above 900 cm^{-1}. Generally speaking, the frequencies are observed at higher values in the oxidized state than in the reduced state. The frequencies around 400 cm^{-1} can be attributed entirely to contributions from either the Cu-Cys or the Cu-His bonds.

Force field parameters for the copper-residue bonds have been extracted from the DFT Hessian matrices, which give frequencies between 100-600 cm^{-1}. The values of these frequencies are shown to depend critically on the in-plane copper-ligand distances, just like was found for the force constants for these in-plane bonds. A decreasing copper-residue

distance leads to higher values for the force constant, and therefore to higher values for the frequencies resulting from the copper force field.

As final check, Molecular Dynamics simulations of 1 ns were performed to check the validity of the copper force field. In these simulations, both the active site and the protein remain stable with root mean square deviations (RMSD) of around 1.0 and 2.0 Å respectively. This stability is confirmed by the copper-residue distances, which show a good agreement with the ones observed in the X-ray data. The distance of the copper from the N_2S plane also showed agreement between simulated and experimental values. This may be a valuable check for the validity of the copper force field, as this distance was not included in any of the bonding interactions of the copper force field.

The vibrational frequencies for the in-plane ligands were obtained in the simulations. The frequencies from the Cu-Cys bond were found primarily between 300 and 400 cm^{-1}, while those for the Cu-His bonds were found either at lower or higher values. Therefore, a very good agreement is observed between the vibrational frequencies observed either by Density Functional Theory, copper force field or Molecular Dynamics simulations.

chapter seven

MAGNETIC INTERACTIONS

in
copper proteins
and
copper complexes

CONTENTS

SUMMARY

The calculation of g-tensors and hyperfine coupling tensors of copper complexes and copper proteins by DFT studies is presented in Chapter 7. It is shown that a procedure is needed in order to get reasonable agreement between computed and experimental g-tensor values. The computed copper hyperfine coupling constants are sometimes in disagreement with the experimental values, but in many cases a good agreement is observed. For the other atoms, especially in wildtype azurin, a good agreement is found between the computed and experimental hyperfine couplings.

7.1 Theory and practice

Magnetic interactions of electrons and molecular systems, in particular copper containing systems

If a free atom carries a magnetic dipole moment μ, it will interact with a magnetic field B, which is described by the Hamiltonian $\mathcal{H}$:[16,282,283]

$$\mathcal{H} = -\mu \cdot B \tag{1}$$

One source of a magnetic moment in an atom is orbital electronic motion that results directly from the angular momentum l of the charge. For an electron moving in an orbit, μ_L is found to have the classical value:

$$\mu_L = (-e/2mc)l \tag{2}$$

where l is its orbital angular momentum, e the charge and m the mass of the electron, and c the velocity of light. This illustrates the generality that μ is proportional to angular momentum, but oppositely directed, for a negatively charged electron. This proportionality is often expressed by the definition of the magnetogyric ratio γ:

$$\mu = \gamma l \tag{3}$$

The electron also spins about its own axis, i.e. it has an intrinsic magnetic moment, but here μ_e is anomalously larger due to spin orbit coupling, by a factor of 2, than for μ_L:

$$\mu_s = 2(-e/2mc)s \tag{4}$$

These orbital and spin contributions are then added to give the total magnetic moment of the atom.

The solution of eq. (1) yields the energy U of the magnetic particle in the field as:

$$U = -\mu B \cos\theta \tag{5}$$

where θ is the angle between the dipole moment μ and the static magnetic field B. Classically, the energy can vary continuously with the orientation of the magnetic moment; quantum mechanically, this is quantized and the angle can take only $2J+1$ orientations, where J is the quantum number for total angular momentum. The allowed projections of J (or μ) along the magnetic field direction are given by m_J, where m_J is the magnetic quantum number with the values:

$$m_J = J, J-1, \ldots, -J \tag{6}$$

If only spin angular momentum arises (as in an atom in a $^2S_{1/2}$ state), m_J becomes:

$$m_s = S, S-1, \ldots, -S \tag{7}$$

with the total electron spin quantum number S. The moment μ_z along the field direction is then (as $l_S = m_S \hbar$):

$$\mu_z = -2(e\hbar/2mc)m_s \tag{8}$$

With the definition of the Bohr magneton μ_B as $e\hbar/2mc$, the (Zeeman) allowed energies for an atom in a magnetic field are found:

$$U_{m_s} = -\mu_z B = 2\mu_B m_s B \tag{9}$$

Thus, in this pure spin case, the familiar equal spacing of 2S+1 energy levels occurs with a separation of $2\mu_B B$.

For a free electron, a small quantum electrodynamics correction requires that this equation be written as:

$$U_{m_s} = g_e \mu_B m_s B \tag{10}$$

where g_e is 2.0023. In this case, $m_J = m_S = \pm 1/2$ and only two Zeeman levels are possible. By applying an oscillating magnetic field of frequency μ_0 perpendicular to B, a transition can be induced from one level to the other, at a particular resonance B_0 (see Figure 7.1.1):

$$h\mu_0 = g_e \mu_B B_0 \tag{11}$$

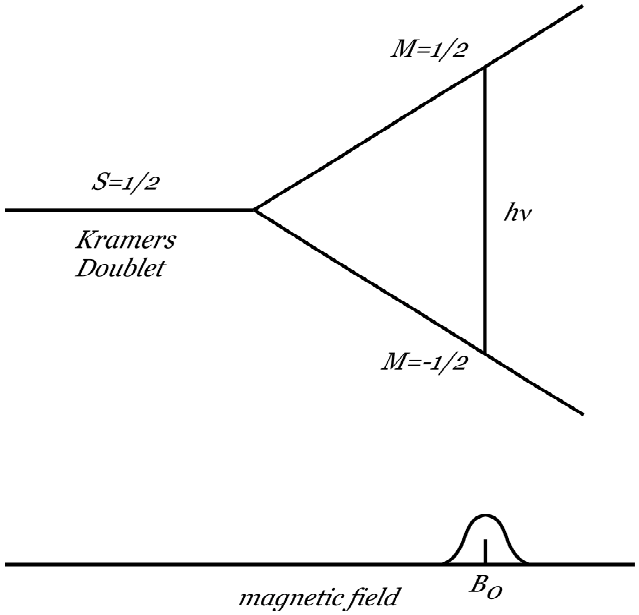


FIGURE 7.1.1. ZEEMAN SPLITTING AND RESONANCE

This is applied in the field of Electron Spin or Paramagnetic Resonance (ESR/EPR), where the strength of the static magnetic field is varied to find the position of the resonance, which results in an EPR spectrum.

Effective spin Hamiltonian

For an electron bound to a copper complex, it is possible to set up an effective spin Hamiltonian in a quantum mechanical formulation as:

$$\hat{\mathcal{H}} = \beta_B \left(\mathbf{B} \cdot \mathbf{g} \cdot \hat{\mathbf{S}} \right) \tag{12}$$

where $\mathcal{H}$ is the operator that, when applied to the wavefunctions of the system, gives the energy eigenvalues. The spin operator vector **S** refers to the effective spin of magnitude 1/2 for the copper complex, and **g** is the g-matrix that describes the interaction of the orbital and spin angular momenta via the spin-orbit coupling. It is symmetric, e.g. $g_{x\,y} = g_{y\,x}$, and the **g**-tensor may be diagonalized by an appropriate choice of the x, y and z directions.

Another important interaction in the spin Hamiltonian is the nuclear hyperfine interaction, which arises from the interaction of the nuclear magnetic moments and the magnetic field at the nuclei generated by the magnetic electrons. As the copper nucleus has a nuclear spin I of 3/2, each electron spin level is split into four components with values of the z component of the nuclear spin quantum number m of –3/2, –1/2, +1/2, +3/2. Thus, the electron resonance line is split into four approximately equally spaced components with a splitting of approximately A, which is the magnitude of the nuclear hyperfine interaction. It has a directional dependence and is therefore more generally written as:

$$\mathcal{H}_{hf} = \mathbf{S} \cdot \mathbf{A} \cdot \mathbf{I} \tag{13}$$

where **I** is the nuclear spin operator vector and **A** the hyperfine interaction matrix, which can be diagonalized by an appropriate choice of axes. In general, these axes are not the same as found for the **g**-tensor, but sometimes they are.

Some more sophisticated additions can be made for the spin Hamiltonian, like the interaction of the nuclear quadrupole moment, but these effects are usually smaller, and may be ignored as a first approximation.

Typical g_Z values for copper complexes are of the order of 2.1-2.6, with a hyperfine splitting of ~0.02 cm^{-1}. As already mentioned in Section 1.2, copper proteins are classified according to three types:

- type 1 copper proteins contain one copper ion, exhibit an unusual EPR spectrum with a hyperfine splitting appreciably smaller than that found for simple copper complexes. They exhibit an intense blue color, g_Z values of 2.2-2.3 and hyperfine splittings around ~0.003-0.008 cm^{-1}.
- type 2 copper proteins exhibit EPR spectra similar to those of simple copper complexes.
- type 3 copper proteins contain a dinuclear copper site and usually, as isolated, are EPR silent, which means that the copper atoms are either in the reduced form, or antiferromagnetically coupled. There are no pronounced features in the optical spectrum visible.

7.2 G-tensors of copper complexes

Prediction of g-tensors of copper molecules by Density Functional Theory calculations

In this section g-tensors are predicted using Density Functional Theory[1] calculations. In the recent literature, it has been reported that it may be difficult to obtain reliable g-tensor values with DFT for complexes containing metal atoms[284-287], although in certain instances a good agreement between calculated and experimental values is observed[288-291]. In this section, a number of copper complexes is used to check the influence on the accuracy of the predicted g-tensors of several issues related to the DFT calculations, like the size of the basis set, (non-) relativistic Hamiltonian, exchange-correlation potential, and the method by which the g-tensor is calculated. Furthermore, the influence of the copper nuclear charge on the results is investigated, as it has been suggested as a possible way to improve the results. The prediction of hyperfine splitting tensors of these molecules is presented in Section 7.3, while the g- and A-tensors for the active sites of copper proteins are described in Section 7.4.

Computational details

Implementations in the ADF code

There are two implementations available in the ADF code[117] for obtaining the g-tensor of a system (with a total spin of 1/2). The first one (ESR), developed by van Lenthe et al.[284,292], uses Gauge Including Atomic Orbitals (GIAO), where the g- and A-tensors are calculated in a Spin-Orbit relativistic calculation using either the Pauli or the more preferred ZORA Hamiltonian[117]. Although the g-tensor can only be calculated in a Spin-Orbit relativistic calculation, the g-tensor of a non-relativistic or scalar relavistic run can still be obtained by using the density from the latter in a Spin-Orbit calculation. The other implementation[291,293] (EPR) is available as an analysis tool, which should be run in a separate program after performing a normal ADF job, which should be either relativistic with the Pauli Hamiltonian or non-relativistic. This implementation does not provide the hyperfine splitting tensor.

Molecule set

A total of eight molecules has been studied[a], for which a comparison of the calculated g- and A-tensors can be made to either experimental data[282,294-296] or other calculated tensors. The geometries of the molecules were taken at their experimental values where available[190,282,297,298], while the remaining unknown coordinates were optimized with the Becke[120]-Perdew[121] potential using a triple zeta valence plus polarization basis in a non-relativistic run of the ADF program.

[a] The molecules are: $CuCl_4^{2-}$ (square planar D_{4h} geometry), $Cu(CO)_3$ (trigonal D_{3h} geometry), CuF_2 (linear), $Cu(H_2O)_6^{2+}$ (D_{2h} symmetry), CuH^+ (linear), $Cu(NO_3)_2$ (D_{2h} symmetry), $CuNO^+$ and CuO (linear)

Results

The principal g-tensor values of a number of copper complexes are given in Table 7.2.1. These values were calculated with Density Functional Theory using the Becke[120]-Perdew[121] exchange-correlation potential in a triple zeta valence plus polarization basis set. The values from both the ESR- and EPR implementations are given where available; for CuO and CuF2 the ESR method was unable to provide the values, as a set of degenerate orbitals was found. This inability has also been found for the EPR implementation, when using the same orbitals from a restricted run. In a subsequent unrestricted run, where the alpha- and beta-orbitals are no longer constrained to be the same, the EPR implementation did provide the g-tensor.

Given in Table 7.2.1 are the ESR values from either a non-relativistic calculation where the ZORA Hamiltonian has been used in the Spin-Orbit calculation for obtaining the g-tensor (nr-zora), a scalar relativistic run using ZORA (sr-zora), a non-relativistic run where the Pauli Hamiltonian was used in the Spin-Orbit run (nr-pauli), a scalar relativistic run using the Pauli Hamiltonian (sr-pauli), and the EPR values from either a non-relativistic run (nr-epr) or a scalar relativistic run using the Pauli Hamiltonian (sr-epr). For the ESR calculations, also the g-tensor from a full Spin-Orbit calculation was obtained, but the difference from the one obtained from the scalar relativistic run is small; the principal values were constantly smaller in the full Spin-Orbit run, but by an amount of less than 0.1.

The deviations from the free electron value (2.0023) of the calculated principal values of the g-tensors are in nearly all cases smaller than observed in experiments. For copperchlorate ($CuCl_4^{2-}$), the deviation from the free electron value is experimentally found to be ~0.22, while the DFT calculations predict a deviation of only ~0.09 (40 %). The same underestimation of the deviation is observed for other molecules like $Cu(NO_3)_2$ (0.12 calculated vs. 0.25 experimental) or $CuNO^+$ (–0.06 calculated vs. –0.11 experimental). The deviations from the free electron value of the principal value from the EPR implementation are similar or smaller than the ones from the ESR implementation. This is only in part due to the different Hamiltonians used (Pauli or ZORA). For comparison, the g-tensors of the Pauli Hamiltonian in the ESR implementation are also given in Table 7.2.1. The Pauli principal values from the ESR implementation differ more from those obtained in the EPR implementation than the values from ZORA/ESR.

In a recent investigation[71] to study the g-tensor of *wildtype* azurin, it was shown that an effective core potential (ECP) description was needed for copper, which included implicitly relativistic corrections. As the inclusion of the ECP on copper improved the calculated g-tensors considerably (from ~2.10 to ~2.23), it was suggested that relativistic corrections may be needed for copper. However, for the copper complexes studied here a rather small difference is observed between the values from a non-relativistic and a (scalar/Spin-Orbit) relativistic calculation, where the principal values of the g-tensor differ by approximately 0.01-0.02 or less. This small influence of relativistic corrections could have been anticipated for a first row transition metal like copper. Although the molecules in this section are smaller than the active site model of *wildtype* azurin, the rather small difference between non-relativistic and relativistic calculations on the copper complexes seems to suggest that relativistic corrections may improve the g-tensors slightly, but might be not the major factor for the underestimation of the g-tensor principal values.

The dependence of the g-tensor on the basis set size can not be ignored completely. In Table 7.2.2, the principal values of the g-tensor from the ESR implementation are given in three basis sets; a small double zeta valence basis (DZV), a medium triple zeta valence plus

polarization basis (TZP) and a large quadruple zeta basis set (QZ). Upon increasing the basis set size from DZV to TZP, the deviation of g-tensor principal values from the free electron value decreases by a small amount in most cases.

TABLE 7.2.1. G-TENSORS FOR BECKE-PERDEW POTENTIAL IN TZP BASIS

	exp.	*nr-zora*[a]	*sr-zora*[b]	*nr-epr*[c]	*sr-epr*[d]	*nr-pauli*[e]	*sr-pauli*[f]
$CuCl_4^{2-}$							
g_x	*2.040*	2.029	2.030	2.032	2.033	2.035	2.036
g_y	*2.040*	2.029	2.030	2.032	2.033	2.035	2.036
g_z	*2.221*	2.087	2.096	2.092	2.096	2.125	2.138
$Cu(CO)_3$							
g_x	*2.0002*	2.0004	2.0011	2.0002	2.0006	1.9990	1.9999
g_y	*2.0002*	2.0004	2.0011	2.0002	2.0006	1.9990	1.9999
g_z	*2.0008*	2.0018	2.0019	2.0009	2.0008	2.0009	2.0009
CuF2							
g_x	*2.601*	-[g]	-[g]	2.244	2.231	-[g]	-[g]
g_y	*2.601*	-[g]	-[g]	2.244	2.231	-[g]	-[g]
g_z	*1.913*	-[g]	-[g]	2.003	2.003	-[g]	-[g]
$Cu(H2O)6^{2+}$							
g_x	*2.095*	2.054	2.060	2.059	2.059	2.064	2.074
g_y	*2.095*	2.075	2.077	2.066	2.066	2.095	2.098
g_z	*2.40*	2.287	2.300	2.214	2.216	2.406	2.426
CuH^+							
g_x	-	2.122	2.147	2.094	2.116	2.179	2.213
g_y	-	2.122	2.147	2.094	2.116	2.179	2.213
g_z	-	1.997	1.995	2.002	2.002	1.990	1.987
$Cu(NO_3)_2$							
g_x	*2.052*	2.025	2.028	2.030	2.031	2.033	2.038
g_y	*2.052*	2.027	2.031	2.032	2.033	2.035	2.041
g_z	*2.249*	2.113	2.126	2.117	2.119	2.162	2.180
$CuNO^+$							
g_x	*2.009*	2.010	2.014	2.006	2.005	2.013	2.018
g_y	*2.009*	2.013	2.017	2.009	2.009	2.017	2.023
g_z	*1.89*	1.939	1.938	1.961	1.965	1.924	1.921
CuO							
g_x	-	-[g]	-[g]	2.049	2.046	-[g]	-[g]
g_y	-	-[g]	-[g]	2.049	2.046	-[g]	-[g]
g_z	-	-[g]	-[g]	2.002	2.002	-[g]	-[g]

a) Non-relativistic density, g-tensor calculated with ZORA Spin-Orbit Hamiltonian
b) Scalar relativistic density, g-tensor calculated with ZORA Spin-Orbit Hamiltonian
c) Non-relativistic unrestricted density, g-tensor calculated with EPR program
d) Pauli Scalar relativistic unrestricted density, g-tensor calculated with EPR program
e) Non-relativistic density, g-tensor calculated with Pauli Spin-Orbit Hamiltonian
f) Scalar relativistic density, g-tensor calculated with Pauli Spin-Orbit Hamiltonian
g) No results available; breakdown due to degenerate orbitals

Increasing the basis set size even more to QZ leads to an increase again, giving in many cases larger deviations from the free electron value than the small DZV basis. For instance for hydrated copper, the scalar relativistic g-tensor z-value goes from 2.32 (DZV) to 2.30 (TZP) to 2.35 (QZ) upon increasing the basis set. For CuH^+ on the other hand, a very small dependence of the g-tensor on the basis is observed. Generally speaking, the best results are

obtained in the large QZ basis, but a rather similar result can be obtained in the DZV or TZP basis.

TABLE 7.2.2. G-TENSORS FOR BECKE-PERDEW POTENTIAL IN THREE BASIS SETS

	exp.	*nr-dzv*	*nr-tzp*	*nr-qz*	*sr-dzv*	*sr-tzp*	*sr-qz*
$CuCl_4^{2-}$							
g_x	*2.040*	2.030	2.029	2.031	2.031	2.030	2.032
g_y	*2.040*	2.030	2.029	2.031	2.031	2.030	2.032
g_z	*2.221*	2.091	2.087	2.098	2.100	2.096	2.110
$Cu(CO)_3$							
g_x	*2.0002*	1.9993	2.0004	2.0016	2.0000	2.0011	2.0025
g_y	*2.0002*	1.9993	2.0004	2.0016	2.0000	2.0011	2.0025
g_z	*2.0008*	2.0019	2.0018	2.0018	2.0019	2.0019	2.0018
$Cu(H2O)6^{2+}$							
7 g_x	*2.095*	2.061	2.054	2.064	2.066	2.060	2.070
g_y	*2.095*	2.076	2.075	2.082	2.079	2.077	2.086
g_z	*2.40*	2.305	2.287	2.334	2.317	2.300	2.349
CuH^+							
g_x	-	2.120	2.122	2.130	2.145	2.147	2.160
g_y	-	2.120	2.122	2.130	2.145	2.147	2.160
g_z	-	1.997	1.997	1.996	1.996	1.995	1.994
$Cu(NO_3)_2$							
g_x	*2.052*	2.024	2.025	2.028	2.027	2.028	2.032
g_y	*2.052*	2.027	2.027	2.030	2.031	2.031	2.034
g_z	*2.249*	2.113	2.113	2.129	2.127	2.126	2.145
$CuNO^+$							
g_x	*2.009*	2.013	2.010	2.014	2.017	2.014	2.019
g_y	*2.009*	2.013	2.013	2.012	2.017	2.017	2.016
g_z	*1.89*	1.910	1.939	1.932	1.909	1.938	1.931

nr) non-relativistic
sr) scalar relativistic (ZORA)
dzv) double zeta valence basis set
tzp) triple zeta valence plus polarization basis set
qz) quadruple zeta valence basis set

The choice of using the Becke-Perdew (BP) exchange-correlation potential could also have an influence on the calculated g-tensor values. Therefore the g-tensors for the molecules were also calculated for four other potentials; the standard local density approximation (LDA), Perdew-Burke-Ernzerhof[228] (PBE) potentials, as well as two potentials that have been specially constructed for an accurate description of excitation energies and (hyper)polarizabilities: statistical averaging of orbital potentials[299] (SAOP) and the gradient regulated asymptotically corrected[300] (GRAC) potential. As these last two potentials seem to give a better description of the (HOMO-LUMO) orbital energies, they might have a marked effect on the g-tensor values also. The computed g-tensor principal values for these five exchange-correlation potentials in the DZV basis set are given in Table 7.2.3. The difference between the principal g-tensor values from the five potentials is of the same order of magnitude as the difference between the different basis sets. For instance for copperchlorate, the computed principal z-value of the g-tensor with either one of the five potentials ranges from 2.096 to 2.111, which is almost exactly the range observed for the

three different basis sets. For hydrated copper, the principal z-value of the g-tensor is computed in the range from 2.317 to 2.344 for the same potentials. This again is almost exactly the range that is observed for the three different basis sets. Therefore, just like the size of the basis set, also the choice of the exchange-correlation is shown to have only a small effect on the computed g-tensors.

Only a small difference is observed between the BP and GRAC results. As the latter is a gradient regulated combination of the BP potential for the inner region and the asymptotically correct LB94 potential for the outer region, this indicates that the outer region is not really that important for a description of the g-tensor. This behavior is observed for all three basis sets used (DZV, TZP and QZ).

TABLE 7.2.3. G-TENSOR PRINCIPAL VALUES FOR XC-POTENTIALS IN DZV BASIS

	exp.	*lda*	*bp*	*pbe*	*saop*	*grac*
$CuCl_4^{2-}$						
g_x	*2.040*	2.030	2.031	2.031	2.032	2.031
g_y	*2.040*	2.030	2.031	2.031	2.032	2.031
g_z	*2.221*	2.096	2.100	2.103	2.107	2.100
$Cu(CO)_3$						
g_x	*2.0002*	2.0004	2.0000	1.9999	1.9991	2.0002
g_y	*2.0002*	2.0004	2.0000	1.9999	1.9991	2.0002
g_z	*2.0008*	2.0019	2.0019	2.0019	2.0019	2.0019
$Cu(H2O)6^{2+}$						
g_x	*2.095*	2.062	2.066	2.066	2.071	2.066
g_y	*2.095*	2.076	2.079	2.080	2.086	2.076
g_z	*2.40*	2.335	2.317	2.325	2.338	2.317
CuH^+						
g_x	-	2.143	2.145	2.144	2.157	2.144
g_y	-	2.143	2.145	2.144	2.157	2.144
g_z	-	1.995	1.996	1.996	1.995	1.995
$Cu(NO_3)_2$						
g_x	*2.052*	2.026	2.027	2.027	2.030	2.027
g_y	*2.052*	2.030	2.031	2.031	2.035	2.031
g_z	*2.249*	2.122	2.127	2.128	2.139	2.127
$CuNO^+$						
g_x	*2.009*	2.010	2.017	2.017	2.017	2.017
g_y	*2.009*	2.012	2.017	2.018	2.024	2.017
g_z	*1.89*	1.884	1.909	1.902	1.882	1.910

Adjustment of the nuclear charge on copper

It has been suggested by Groenen and Solomon[a] that a possible way of improving the computed g-tensor values is by changing the nuclear charge of copper. The assumption is that the copper 3d orbitals are too low in energy to mix properly with the ligand orbitals, thereby producing a too low contribution of copper orbitals in the singly occupied molecular orbital (SOMO). The adjustment of the copper nuclear charge would effectively push the 3d-orbitals on copper upwards, leading to larger contributions of copper orbitals in the highest occupied molecular orbitals (including the SOMO), which would lead to increased z-values of

[a] E.J.J. Groenen, E.I. Solomon, personal communication

the g-tensor. Alternatively, this procedure would also lead to improved excitation energies (see Chapter 8).

For the same set of molecules, the nuclear charge on copper has been adjusted in steps of 0.2 electron, shifting it downwards from 29.0 to 28.2. The computed principal g-tensor values in the DZV basis set using the Becke-Perdew xc-potential are given in Table 7.2.4. As expected, the principal values of the g-tensor increase with decreasing copper nuclear charge. This pattern is not only observed for the z-values of the g-tensor, but also for the x- and y-values.

For CuH^+, some difficulties arise when adjusting the copper nuclear charge. Just like found for CuF_2 and CuO, some molecular orbitals become degenerate with certain copper nuclear charges, leading to erroneous g-tensors.

TABLE 7.2.4. G-TENSOR VALUES AS FUNCTION OF COPPER NUCLEAR CHARGE (Z)

	exp.	*Z=29.0*	*Z=28.8*	*Z=28.6*	*Z=28.4*	*Z=28.2*
$CuCl_4^{2-}$						
g_x	*2.040*	2.031	2.039	2.053	2.070	2.084
g_y	*2.040*	2.031	2.039	2.053	2.070	2.084
g_z	*2.221*	2.100	2.141	2.187	2.232	2.271
$Cu(CO)_3$						
g_x	*2.0002*	2.0000	2.0015	2.0030	2.0044	2.0058
g_y	*2.0002*	2.0000	2.0015	2.0030	2.0044	2.0058
g_z	*2.0008*	2.0019	2.0019	2.0019	2.0019	2.0019
$Cu(H_2O)_6^{2-}$						
g_x	*2.095*	2.066	2.078	2.096	2.104	2.098
g_y	*2.095*	2.079	2.091	2.108	2.110	2.110
g_z	*2.40*	2.317	2.347	2.396	2.417	2.426
CuH^+						
g_x	-	2.145	2.218	2.283[a]	2.195	2.316[b]
g_y	-	2.145	2.218	2.283[a]	2.195	2.316[b]
g_z	-	1.996	1.990	1.983[a]	1.993	1.983[b]
$Cu(NO_3)_2$						
g_x	*2.052*	2.027	2.037	2.047	2.051	2.045
g_y	*2.052*	2.031	2.044	2.059	2.073	2.083
g_z	*2.249*	2.127	2.161	2.191	2.215	2.231
$CuNO^+$						
g_x	*2.009*	2.017	2.023	2.026	2.028	2.029
g_y	*2.009*	2.017	2.026	2.039	2.056	2.078
g_z	*1.89*	1.909	1.912	1.918	1.925	1.935

a) g-tensor from non-relativistic density

b) value obtained in TZP basis

Based on the g_z-values with the largest deviation from the free electron value (copper chlorate, hydrated copper and copper dinitrate), the "optimal" value of the copper nuclear charge for the prediction of good g-tensors would then be around 28.4-28.6, depending on which basis set one uses. For TZP a value of 28.4 seems more appropriate, for QZ a value of 28.6 while it is somewhere inbetween for the DZV basis. These "optimal" values seem to be rather insensitive to the xc-potential used. In the following section, the prediction of

hyperfine splitting A-tensors is discussed for the same molecules, while the application to the active sites of copper proteins is presented in Section 7.4.

Conclusions

The computed g-tensors of copper complexes using Density Functional Theory are, without using special procedures to improve the results, rather disappointing. Generally speaking, only half of the experimental deviation from the free electron value is recovered by the calculations. Including scalar or spin-orbit relativistic effects or increasing the basis set size have only a marginal, yet improving, effect.

The computed values can be improved by artificially lowering the copper nuclear charge, which has the effect that effectively the copper d-orbitals are shifted upwards and the contributions of these in the singly occupied highest occupied molecular orbital are increased. This effectively leads to larger g-tensor values and therefore to a better agreement with experimental data.

7.3 *Hyperfine splittings*

Prediction of the A-tensor of copper complexes

In Section 7.2, the prediction of the g-tensor of copper complexes has been discussed, while in this section the calculation of hyperfine splitting (A-) tensors is presented. Unlike the g-tensors for which one value is obtained for the whole quantum system, the A-tensors are obtained per atom. The tensors are best obtained in an unrestricted calculation, e.g. a calculation where the Kohn-Sham orbitals for the alpha- and beta-electrons, and therefore the alpha- and beta-spin density, are not constrained to be the same.

TABLE 7.3.1. XC-POTENTIAL DEPENDENCE OF HYPERFINE SPLITTINGS[a] (MHZ)

	exp.[282,301]	*LDA*	*BP*	*PBE*	*SAOP*	*GRAC*
$CuCl_4^{2-}$						
Cu	-234	-152	-183	-179	-316	-188
Cl	-	24	24	24	9	24
$Cu(CO)_3$						
Cu	71	-11.3	-0.8	-0.5	345	20.4
C	-18.7	-19.4	-20.9	-17.3	6	-14.0
O	-11.2	4.6	0.7	-1.0	-70	-6.2
CuF_2						
Cu	2038[b]	-273	-229	-	-	-227
F	240[b]	-123	-77	-	-	-82
$Cu(H_2O)_6^{2+}$						
Cu	-	-175	-190	-189	-531	-195
O_{ax}	-	3	4	4	3	2
O_{eq}	-	-101	-97	-97	-80	-98
CuH^+						
Cu	-	3223	2972	-	3519	2964
H	-	517	532	-	860	532
$CU(NO_3)_2$						
Cu	-223	-209	-239	-233	-335	-245
O	-	-2	-6	-8	26	-7
N	-	-4	-3	-3	7	-3
$CuNO^+$						
Cu	570	2548	2470	2412	2366	2448
N	-	42	42	45	-17	43
O	-	18	5	-1	72	2
CuO						
Cu	-484	-807	-718	-663	-1138	-645
O	-	52	37	24	307	37

a) isotropic value of hyperfine splitting tensor in DZV basis set

b) absolute value

In Section 7.2, it was shown that the choice of the exchange-correlation potential had little effect on the calculated g-tensor values. Whether this is true also for the hyperfine

splitting tensor, is examined first. The computed isotropic values of the A-tensor for the same five xc-potentials (LDA[119], BP[120,121], PBE[228], SAOP[299] and GRAC[300]) as used in Section 7.2 are given in Table 7.3.1.

Apart from the SAOP potential, only relatively small differences are observed between the five xc-potentials. For instance for copper chlorate, the isotropic hyperfine splitting value of ranges from –152 to –188 MHz copper and is 24 MHz on chloride for all potentials; for hydrated copper the isotropic hyperfine splitting ranges from –175 to –195 MHz on copper and from –97 to –101 MHz on the equatorial oxygen. For CuH^+ some larger differences are observed on copper (2964 to 3223 MHz), but for hydrogen the value remains rather constant at 517-532 MHz. Still, these larger differences are still small compared to the difference between the molecules, and may therefore not be significant.

In most cases, the order of magnitude of the isotropic hyperfine splittings is correct in comparison with the experimental value, apart from the copper hyperfine splitting in CuF_2 and $CuNO^+$. In the former, the computed values are too small by a factor of 7-8, while in the latter, the computed values are too large by factor of 4-5.

TABLE 7.3.2. BASIS SET DEPENDENCE OF HYPERFINE SPLITTINGS[a] (MHZ)

	exp. [282,301]	*DZV*	*TZP*	*QZ*
$CuCl_4^{2-}$				
Cu	-234	-183	-183	-248
Cl	-	24	27	29
$Cu(CO)_3$				
Cu	71	-0.8	-17.8	-8.0
C	-18.7	-20.9	-14.7	-16.4
O	-11.2	0.7	-5.4	-3.0
CuF_2				
Cu	2038	-229	-213	-282
F	240	-77	-81	49
$Cu(H_2O)_6^{2+}$				
Cu	-	-190	-182	-
O_{ax}	-	4	3	-
O_{eq}	-	-97	-96	-
CuH^+				
Cu	-	2972	3016	3237
H	-	532	469	477
$CU(NO_3)_2$				
Cu	223	-239	-238	-316
O	-	-6	-21	-17
N	-	-3	-4	-3
$CuNO^+$				
Cu	570	2470	2940	3170
N	-	42	48	44
O	-	5	-21	-11
CuO				
Cu	-484	-718	-661	-677
O	-	37	-25	-6

a) isotropic value of hyperfine splitting tensor using Becke-Perdew potential

The influence of the choice of basis set on the computed values is checked using the Becke-Perdew potential in the same three basis sets that have been used in Section 7.2 (small DZV, medium TZP and large QZ basis sets). The computed isotropic hyperfine splitting values are given in Table 7.3.2. It is observed that the basis set size has some effect on the computed isotropic hyperfine splitting values, but not the dramatic changes needed to get good values compared to the experimental results. This could have been anticipated, as the influence of the basis set on the computed g-tensor values was shown to be small. For $CuNO^+$, the copper hyperfine splitting increases from 2470 MHz in the DZV basis to 3170 MHz in the QZ basis set, and thereby reduces the agreement with the experimental value even more. On the other hand for copper chlorate, the difference between the computed and experimental value decreases to only 10 MHz.

TABLE 7.3.3. HYPERFINE SPLITTINGS[a] (MHZ) AS FUNCTION OF NUCLEAR CHARGE Z

atom	*exp.*	*Z=29.0*	*Z=28.8*	*Z=28.6*	*Z=28.4*	*Z=28.2*
$CuCl_4^{2-}$						
Cu	-234	-183	-186	-196	-217	-247
Cl	-	24	29	33	35	36
$Cu(CO)_3$						
Cu	71	-0.8	-2.1	-2.8	-3.1	-3.3
C	-18.7	-20.9	-20.5	-20.1	-19.6	-19.1
O	-11.2	0.7	0.7	0.6	0.6	0.6
CuF_2						
Cu	2038	-229	-237	-249	-263	-273
F	240	-77	-55	-38	-28	-26
$Cu(H_2O)_6^{2+}$						
Cu	-	-190	-191	-	-	-326
O_{ax}	-	4	4	-	-	43
O_{eq}	-	-97	-89	-	-	-94
CuH^+						
Cu	-	2972	334	138	25	-18
H	-	532	205	149	88	31
$CU(NO_3)_2$						
Cu	-223	-239	-241	-250	-267	-290
O	-	-6	-13	-19	-24	-28
N	-	-3	-3	-2	-2	-3
$CuNO^+$						
Cu	570	2470	2557	2696	2846	2961
N	-	42	38	37	36	37
O	-	5	5	4	3	2
CuO						
Cu	-484	-718	-656	-616	-567	-534
O	-	37	32	27	22	19

a) isotropic hyperfine splitting tensor value in DZV basis with Becke-Perdew potential

In Section 7.2, it was observed that a special procedure was needed to get reasonable agreement between computed and experimental g-tensor values. This procedure consisted of lowering the nuclear charge on copper, which would effectively shift the d-orbitals of copper upwards. This leads to larger contributions of these orbitals in the singly occupied

molecular orbital of the unpaired electron, and therefore to increased g-tensor values. The influence of this procedure on the computed isotropic hyperfine splitting values is reported in Table 7.3.3. The computed isotropic hyperfine splitting values show different trends for different molecules; for some, the values get more negative, for others they get more positive. The most striking example in this case is CuH^+, in which the copper hyperfine splitting changes from 2972 MHz to –18 MHz, while for hydrogen it changes from 532 MHz to 31 MHz.

The drastic improvement of the computed g-tensor values is however not accompanied by an evenly drastic improvement of the hyperfine splittings. The molecules for which the hyperfine splitting values were already properly described, remain to be so when using this procedure, while for the molecules for which a discrepancy exists between the experimental and computed values, also remain being so. Just like was observed for using the larger QZ basis set, using the special procedure, that helped to improve the computed g-tensor, reduces the agreement between the computed and experimental value of for instance $CuNO^+$.

Conclusions

The hyperfine splitting (A-) tensor of the atoms in copper complexes is computed using Density Functional Theory calculations. The computed values are in most cases of the right order of magnitude in comparison with the experimental values, but in some cases a large discrepancy exists between the two. Although the computed values may vary to some extent by using different basis sets and/or exchange-correlation potentials, the changes in computed values are relatively small. The molecules for which a reasonable agreement between computed and experimental values was observed, remain being properly described, while the molecules that showed a large discrepancy between experimental and computed values, remain to do so.

The special procedure used in Section 7.2 to give a drastic improvement of the computed g-tensor values, has a small effect on the computed hyperfine splitting values, about the same order of magnitude observed when using a different basis set and/or exchange-correlation potential. Therefore, although it has a major impact on the g-tensor values, the influence on the hyperfine splitting values is relatively small, apart from the CuH^+ molecule, in which the copper hyperfine splitting value changes from 2972 MHz to –18 MHz, and the hydrogen value from 532 MHz to 31 MHz.

7.4 *Copper proteins*

Prediction of g-tensors and hyperfine splittings of (models of) the active sites of copper proteins

Until now in this chapter the computation of g-tensors and hyperfine splitting values has been limited to copper complexes, to check whether the computed values are in general agreement with the experimental ones, and to see how they can be improved. In this section, the computation of g-tensors and hyperfine splittings of (active sites of) copper proteins will be presented and discussed. Based on the experience in the previous sections with the choice of the basis set and/or exchange-correlation potential, which was shown to have a small effect, the Becke[120]-Perdew[121] potential is used in a DZV basis set. The potential has been used throughout the rest of the thesis as well, while the calculations will be finished faster when using this smaller basis set.

Several models for the active sites of copper proteins can be thought of, the simplest being simply a copper and a cysteinate (*model 1*). As this ligand is present in all type 1 copper proteins, and its presence is thought to be responsible to a great extent for the strong blue color and the characteristic EPR-features, it might be interesting to see how the computed EPR-parameters compare with larger active site models and experiments. The second model consists apart from the cysteinate also of the other two strong histidine ligands (*model 2*). This model has also been used in an investigation by van Gastel et al.[71] to study the EPR-parameters for wildtype azurin. They observed that it was necessary to use an effective core potential (ECP) on copper that includes relativistic corrections. As the use of the ECP improved the computed g-tensors considerably, it was concluded that relativistic effects should be included to get a good description of the g-tensor. However, as was shown in Section 7.2, the inclusion of (scalar or spin-orbit) relativistic effects in DFT calculations has only a limited effect on the computed g-tensor. The values do shift upwards, but only by 0.01 or 0.02. The major impact of using the ECP is to shift the copper d-orbitals upwards, just like observed for the lowering of the copper nuclear charge.

The first two models can more or less be used as typical models for type 1 copper proteins. More realistic active site models are obtained if also the axial ligands are included in the calculations (*model 3*). In that case, one should be able to distinguish different proteins like *wildtype* azurin, M121Q or M121H azurin. In all these models the amino acid residues are cut off at the carbon-alpha position with the backbone connections replaced by hydrogens.

G-tensors

The computed principal g-tensor values for the models are listed in Table 7.4.1, again also using the suggested procedure (involving a lowering of the copper nuclear charge) to improve the computed values. In Section 7.2, it was shown that a value around 28.4-28.6 for the copper nuclear charge seemed appropriate to calculate reasonably accurate g-tensors for copper complexes. Assuming that the same trend is observed in copper proteins, the g-tensors were calculated using the normal charge (29.0) as well as the ones that performed best for the copper complexes (28.4 and 28.6).

Just like was found for the copper complexes, using a lower copper nuclear charge shifts the computed g-tensor values upwards, to values that compare well with experimental values. This pattern is observed for all models, except the most simple one (*model 1*). In that case, the lowering of the copper nuclear charge has a negligible effect, and seems to even lower the computed g-tensor values. Apparently, the interaction with the histidines is necessary to get a more balanced distribution of the unpaired electron.

TABLE 7.4.1. G-TENSORS OF ACTIVE SITE MODELS (WITH CU NUCLEAR CHARGE Z)

	exp.	*Z=29.0*	*Z=28.6*	*Z=28.4*
model 1:				
g_x	-	2.031	2.021	2.002
g_y	-	2.095	2.046	2.019
g_z	-	2.119	2.093	2.042
model 2:				
g_x	-	2.028	2.059	2.081
g_y	-	2.077	2.092	2.109
g_z	-	2.111	2.244	2.310
model 3:				
g_x *wt*	2.039[74,302-304]	2.029	2.088	2.124
g_y *wt*	2.057[74,302-304]	2.072	2.111	2.143
g_z *wt*	2.273[74,302-304]	2.126	2.278	2.338
g_x *m121q*	2.028[38]	2.033	2.063	2.085
g_y *m121q*	2.083[38]	2.072	2.102	2.121
g_z *m121q*	2.288[38]	2.108	2.199	2.238
g_x *m121h*	2.051[23]	1.976	1.881	1.806
g_y *m121h*	2.056[23]	1.999	1.938	1.865
g_z *m121h*	2.265[23]	2.081	2.172	2.296
g_x *m121e_p*[a]	2.03[77]	2.027	1.920	2.002
g_y *m121e_p*[a]	2.085[77]	2.063	2.028	2.016
g_z *m121e_p*[a]	2.30[77]	2.120	2.219	2.070
g_x *m121e_dp*[b]	2.06[77]	-	-	2.017
g_y *m121e_dp*[b]	2.06[77]	-	-	2.137
g_z *m121e_dp*[b]	2.30[77]	-	-	2.200

a) m121e_p: Glu121 protonated
b) m121e_dp: Glu121 deprotonated

A different behavior is observed for wildtype azurin and M121Q azurin regarding the g_z value; the values are found to be higher for the former than for the latter. For wildtype azurin, a good agreement between the computed g_z value (using a copper nuclear charge of 28.6) and the experimental value is found. Although the g_x and g_y values are somewhat larger than found experimentally, the computed difference between the two (0.023) is again in good agreement with the experimental value of 0.018. The M121Q mutant shows much more rhombic values, with a computed difference between the g_x and g_y values (0.039) that is almost twice as large as for wildtype azurin. This is again in good agreement with the

experimental data (0.055) that indicate a rhombic EPR spectrum, although the g_z value is too small compared to the experimental value.

For the M121H mutant, unexpected results are obtained; the g_x and g_y values are in all cases even below the free electron value of 2.0023. The same is observed for the M121E mutant, for which g-tensor values are obtained at two different pH values (at pH=4 and pH=8); at low pH, the Glu121 residue is protonated at the $O_{\varepsilon 2}$ position, while it is deprotonated at high pH. Preliminary calculations using the TZP basis set, with the same result, seem to indicate that this unexpected behavior is not due to the relatively small basis set used.

Hyperfine splitting tensors

The hyperfine splitting tensors have been computed, just like in the previous section, in a spin-unrestricted run. The computed isotropic hyperfine splittings are given in Table 7.4.2 for copper and some other atoms.

TABLE 7.4.2. ISOTROPIC HYPERFINE SPLITTING[a] (MHZ) OF ACTIVE SITE MODELS

	Cu	O_{45}	$N_{46,\delta}$	$N_{46,\varepsilon}$	S_{112}	$N_{117,\delta}$	$N_{117,\varepsilon}$	X_{121}[b]
model 1	-276.7	-	-	-	-5.9	-	-	-
model 2	-167.8	-	14.4	0.8	-1.3	20.3	1.4	-
model 3:								
wt	-157.5	0.1	16.1	0.9	-0.3	22.4	1.5	-2.1
m121q	-89.7	0.3	17.6	1.3	-1.3	28.7	1.7	0.7
m121h	-59.9	0.0	13.2	0.3	-1.4	30.4	1.9	10.2
m121e_p[c]	-119.8	1.0	19.4	1.1	-0.4	26.1	1.8	-2.9
m121e_dp[d]	-88.8	0.7	22.6	1.5	-0.4	26.5	2.1	-11.1

a) computed values in DZV basis with copper nuclear charge of 28.6;
b) X=S, O, N
c) m121e_p: Glu121 protonated
d) m121e_dp: Glu121 deprotonated

Compared to model 1, the copper hyperfine splitting decreases upon inclusion of the histidine residues in model 2, as well as the hyperfine splitting on sulphur. Including also the axial ligands in model 3 (*wt*) even decreases the values even some more. Still, the values for copper is too large by a factor of three compared to experimental data. However, the principal values of the A-tensor are, relatively speaking, in good agreement with experimental data, which show in both cases a large value in the z-direction and almost negligible values in the x- and y-direction.

The computed hyperfine splittings of the nitrogens in the histidine residues in the *wt* structure agree very well with the experimental values of 18.1 (N_δ-His46), 0.9 (N_ε-His46), 25.1 (N_δ-His117) and 1.3 MHz (N_ε-His117). Also the hydrogen hyperfine couplings are in good agreement (see Table 7.4.3). The largest values are found for the H_β protons of Cys112, with computed values of 31.5 and 28.2 MHz that agree very well with the experimental values of 27 and 28 MHz. Also the hyperfine coupling of the H_α of Cys112 is well described (–0.4 MHz

for both the experimental and computed value). The hyperfine couplings of the protons in the histidines are reasonably well described; there is either a good agreement (for instance for $H_{\varepsilon 1/\varepsilon 2}$-His46, $H_{\delta 2}$-His46) or agreement within the limit of the accuracy of both methods.

The isotropic hyperfine couplings of copper in the M121Q, M121H and M121E mutants is considerably lower than in *wildtype* azurin. The hyperfine coupling of the other atoms doesn't change significantly apart from the atom in residue 121 in the M121H and *deprotonated* M121E mutants, where the absolute value increases to roughly 10 MHz.

TABLE 7.4.3. COMPARISON OF PROTON HYPERFINE COUPLINGS (MHZ) OF PROTONS IN ACTIVE SITE OF *WILDTYPE* AZURIN

proton	*experimental*	*computed*
H δ2 His46	1.49	0.79
H ε1 His46	1.06/1.48	1.32
H ε2 His46	0.56	0.60
H β1 Cys112	28/27	31.5
H β2 Cys112	27/28	28.2
H α Cys112	-0.38	-0.44
H δ2 His117	1.61	2.00
H ε1 His117	1.45/1.02	0.79
H ε2 His117	-	1.06

For the Met121Glu mutant, two different sets of hyperfine tensors are observed on copper, one at low pH where Glu121 is protonated and one at high pH where Glu121 is deprotonated. Even though the principal values of the g-tensor change by only a relatively small amount, the principal values of the hyperfine coupling tensor change dramatically. In the protonated form the A_X value (182 MHz) is larger than the A_Z value (70 MHz), while in the deprotonated form the A_X value (28 MHz) is smaller than the A_Z value (232 MHz). Even though the computed values are too large in the absolute sense, the same pattern is observed with values of 453/109 MHz (A_X/A_Z) in the protonated form and 214/427 MHz (A_X/A_Z) in the deprotonated form.

Conclusions

The g-tensor and atomic hyperfine couplings were computed for several active site models of copper proteins and the results compared with experimental data. For the g-tensor a reasonable agreement is observed for *wildtype* azurin and the M121Q mutant, especially when looking which type of spectra (rhombic/axial) is observed, but peculiar results are obtained for the M121H and M121E mutants. Further investigation is needed for these mutants, which is currently being done.

The hyperfine coupling of copper is too large by a factor of three, but the relative principal values of the A-tensor are well described, with a large value in the z-direction and almost negligible values in the x- and y-direction. The hyperfine couplings of the nitrogen

atoms in the histidine residues show a good agreement between the experimental and computed values, with larger values observed in the His117 residue. Also the relative values of the N_δ vs. the N_ε atoms is well described. Also for the protons in the histidine and cysteine residues is a good agreement between experimental and computed values observed for *wildtype* azurin.

In the Met121Glu mutant of azurin two sets of experimental data are available, corresponding to Glu121 being either protonated or deprotonated; the principal values of the copper hyperfine tensor change dramatically in relative values. In the protonated form, the hyperfine coupling in the x-direction is largest, while in the deprotonated form the hyperfine coupling in the z-direction is largest. Apart from the fact that the computed values are too large by a factor of roughly three, the same change in relative values is observed.

chapter eight

UV/VIS SPECTRA *of copper complexes*

8

CONTENTS

SUMMARY

This chapter deals with the UV/VIS spectra, or excitation energies, of copper proteins. As Time-Dependent Density Functional Theory (TD-DFT) is formally not applicable to open shell systems, use is made of the semi-empirical CNDO/1 and INDO/1 approaches. TD-DFT results for the reduced state of three copper proteins is used to calibrate the CNDO/INDO copper parameters, which are subsequently tested for the excitation energies in the oxidized state of these proteins. Also the influence of the presence of the protein on the computed excitation energies is checked.

8.1 *Calibration*

A direct comparison of Density Functional Theory and semi-empirical excitation energies of the reduced state of copper proteins

In recent years Density Functional Theory (DFT)[1] has been successfully applied to the computation of excitation energies and polarizabilities of (metallo-)organic molecules, by using Time-Dependent DFT (TD-DFT). It was shown to be an efficient and reliable method to compute these response properties, that compare well with experimental data. Unfortunately, TD-DFT is in principle possible only for closed-shell systems. As the most important of the two states of copper proteins (oxidized state) has an unpaired electron, and therefore is an open shell system with a doublet ground state, formally TD-DFT can not be used for the computation of the excitation energies of an active site of copper proteins in the oxidized state.

Alternatively, the excitation energies can be obtained by semi-empirical calculations using the INDO or CNDO Hamiltonian[a], which have been successfully used for the UV/VIS spectra of (metallo)organic molecules. As the excitation energies are then obtained in a Configuration Interaction calculation, no problem arises for an open shell ground state and also the oxidized state of copper proteins can be treated with it. A remaining issue that needs to be solved before one can use it straightforwardly for metalloproteins, is to check whether the standard INDO or CNDO parameters give a reasonable description for the system. Therefore the excitation energies of the reduced state of the copper proteins will be computed by using both TD-DFT and the semi-empirical approach, and the copper parameters needed in the latter will be adjusted if necessary.

An important aspect in the TD-DFT investigations was the recognition that standard exchange-correlation potentials do not give an accurate description of the energy levels of the molecular orbitals. Therefore specially designed xc-potentials[216,299,300] have been constructed that improve the computed excitation energies (as well as other properties like the molecular polarizability[210]) enormously. One of the first of these potentials is the van Leeuwen-Baerends potential (LB94)[216] that shows the correct asymptotic behavior, which is needed to obtain good polarizability values. However even though the outer region is described well, the inner region is less well described. Therefore a combination of a potential giving a good description of the inner region (Becke[120]-Perdew[121]) and a potential giving a good description of the outer region (LB94) has been tried where the weight of each potential at a certain point is gradient regulated in order to get a smooth transition from the inner to the outer region. One additional feature of this gradient regulated asymptotically correct (GRAC)[300] potential is that it needs the ionization potential of the molecule as input parameter. For molecules for which this property is known, one can simply use the experimental value, otherwise one has to use a computed value. In the latter case, it is recommended to use the value obtained by using the SAOP (Statistical Averaging of model Orbital Potentials) potential[299], as this usually provides ionization potentials that compare well with experimental values.

The energy level of the Highest Occupied Molecular Orbital (HOMO) using a triple zeta valence basis set plus polarization functions (TZP) for the reduced state of the active site of

[a] The abbreviations CNDO and INDO stand for the CNDO/1 and INDO/1 Hamiltonians

three copper proteins with three exchange-correlation potentials is reported in Table 8.1.1. For *wildtype* azurin[a] the orbital energy was computed with the Becke-Perdew (BP) potential, the SAOP potential and the GRAC potential (where the ionization potential obtained in the SAOP calculation has been used as input parameter), both in a non-relativistic and scalar relativistic[b] run. The influence of including relativistic corrections is negligible; the energy level of the HOMO changes by roughly 0.001 Hartree (~0.04 eV), which is probably well below the accuracy of the computed energies. The ionization potential (corresponding to minus the orbital energy of the HOMO) is increased considerably by using the SAOP potential, which has been seen also for small molecules.

TABLE 8.1.1. ORBITAL ENERGY OF HIGHEST OCCUPIED MOLECULAR ORBITAL (A.U.)

	m121h	*m121q*	*wildtype*
BP/NR[a]	-[c]	-[c]	-0.0823
BP/SR[b]	-[c]	-[c]	-0.0809
SAOP/NR[a]	-0.2075	-0.2095	-0.2219
SAOP/SR[b]	-[c]	-[c]	-0.2214
GRAC/NR[a]	-0.0734	-0.0739	-0.0824
GRAC/SR[b]	-[c]	-[c]	-0.0810

a) non-relativistic results
b) scalar relativistic results using ZORA approach;
c) not performed due to experiences with *wildtype* results (see text)

As the influence of relativistic corrections is small, and the orbital energy of the HOMO in the Becke-Perdew calculation too high (too less negative), for the M121H and M121Q mutants of azurin only the SAOP potential in a non-relativistic run was used to compute the HOMO energy level, which is found to lie 0.3-0.4 eV higher than for wildtype eV.

Using the ionization potentials obtained with the SAOP potential as input for the GRAC potential, the fifteen lowest excitation energies for the reduced state of the proteins have been computed using TD-DFT. The values and the corresponding oscillator strengths are given in Table 8.1.2 for *wildtype*, M121H and M121Q azurin. Just as the ionization potentials of the M121H/M121Q mutants were lower than the one of *wildtype* azurin, also the lowest excitation energy is found at lower energy. Apart from the oscillator strengths, a remarkable similarity between the M121H and M121Q results is obtained.

For *wildtype* azurin the HOMO consists mainly of the p_y orbital of the Cys112 sulphur (49 %) and the d_{xy} orbital of copper (37 %). These numbers are roughly the same as the ones obtained in Section 6.3. In M121H azurin, it is no longer only the p_y orbital of sulphur that contributes (46 %), but also the p_x (3 %) and p_z (2 %) orbitals. The same goes for copper, where next to d_{xy} (26 %) also $d_{x^2-y^2}$ (7 %) contributes. This is also true for M121Q azurin, but to a lesser extent; the HOMO now consists primarily of the p_y orbital of the Cys112 sulphur (49 %) and the d_{xy} (34 %) and $d_{x^2-y^2}$ (2 %) orbitals of copper.

[a] The active site of *wildtype*, M121H and M121Q consists in this chapter of the five groups coordinating to copper (Gly45, His46, Cys112, His117, Met/His/Gln121), which are cut-off at the C-alpha position
[b] Scalar relativistic ZORA calculation[117,224]

TABLE 8.1.2. EXCITATION ENERGIES (EV) AND OSCILLATOR STRENGTHS

m121h		*m121q*		*wildtype*	
GRAC		*GRAC*		*GRAC*	
energy	*osc.str.*	*energy*	*osc.str.*	*energy*	*osc.str.*
1.276	0.0055	1.252	0.0120	1.562	0.0079
1.396	0.0007	1.334	0.0002	1.691	0.0031
1.608	0.0038	1.755	0.0025	1.934	0.0169
1.652	0.0100	1.805	0.0051	2.043	0.0003
1.773	0.0024	1.844	0.0164	2.152	0.0026
1.827	0.0038	1.954	0.0016	2.238	0.0132
1.903	0.0208	1.976	0.0036	2.280	0.0007
1.927	0.0059	2.080	0.0006	2.346	0.0033
1.984	0.0132	2.113	0.0102	2.358	0.0024
2.093	0.0000	2.208	0.0008	2.455	0.0028
2.197	0.0030	2.342	0.0019	2.584	0.0080
2.260	0.0031	2.353	0.0018	2.699	0.0041
2.265	0.0041	2.511	0.0042	2.709	0.0027
2.292	0.0002	2.519	0.0015	2.776	0.0021
2.322	0.0029	2.541	0.0129	2.792	0.0013
CNDO[a]		*CNDO*[a]		*CNDO*[a]	
energy	*osc.str.*	*energy*	*osc.str.*	*energy*	*osc.str.*
1.45	0.0012	1.19	0.0002	1.87	0.0005
1.65	0.0003	1.50	0.0005	1.96	0.0001
1.72	0.0001	1.66	0.0007	2.06	0.0029
1.76	0.0003	1.75	0.0002	2.08	0.0008
1.91	0.0319	1.89	0.0379	2.14	0.0041
2.02	0.0020	1.96	0.0103	2.16	0.0003
2.13	0.0050	2.12	0.0114	2.20	0.0247
2.20	0.0011	2.20	0.0294	2.30	0.0103
2.24	0.0002	2.29	0.0036	2.35	0.0025
2.31	0.0081	2.37	0.0047	2.44	0.0173
INDO[b]		*INDO*[b]		*INDO*[b]	
energy	*osc.str.*	*energy*	*osc.str.*	*energy*	*osc.str.*
1.40	0.0007	1.32	0.0004	1.69	0.0004
1.63	0.0003	1.60	0.0004	1.87	0.0029
1.68	0.0002	1.67	0.0004	1.88	0.0001
1.75	0.0000	1.77	0.0001	1.92	0.0005
1.93	0.0175	2.02	0.0096	1.96	0.0001
1.96	0.0205	2.08	0.0545	2.04	0.0422
2.09	0.0070	2.21	0.0103	2.11	0.0018
2.13	0.0013	2.31	0.0059	2.13	0.0019
2.15	0.0005	2.36	0.0055	2.17	0.0050
2.24	0.0233	2.39	0.0066	2.40	0.0055

a) using “optimized” CNDO copper parameters (see text)
b) using “optimized” INDO copper parameters (see text)

In the semi-empirical (CNDO or INDO) approach, standard atomic parameters are used to calculate the one- and two electron integrals, where use has been made of empirical parameters like ionization potentials of the atoms. For organic molecules these parameters

are well calibrated and usually a good agreement between experimental and computed excitation energies are observed.

For systems containing metal atoms, a few different parameterizations are present, which were optimized for particular systems; one can be used for copper complexes with strong interactions to copper[305,306], while another set (Clack's values)[307,308] was constructed for systems with weaker bonds to copper. The first set can not be used for copper proteins in the reduced state as it gives too high excitation energies: the lowest energies are found at respectively 3.07/3.16 eV (M121H), 2.66/2.79 eV (M121Q) and 3.76/3.79 eV (*wildtype*) for the CNDO/INDO Hamiltonians, which lie some 1.4-2.1 eV above the lowest DFT excitation. For the lowest ten energies, the average absolute deviation from the DFT values is 2.00 eV for CNDO and 2.08 for INDO, with a corresponding deviation for the oscillator strengths of 0.018 (CNDO) and 0.017 (INDO).

Using Clack's values[307,308] on the contrary gives excitation energies that are lower than the DFT energies, respectively 1.12/0.76 eV (M121H), 0.93/0.63 eV (M121Q) and 1.36/1.02 eV (*wildtype*) for CNDO/INDO. However, the deviations are smaller, 0.23 (CNDO) and 0.45 (INDO) eV for the energy and 0.013/0.014 for the oscillator strength.

The copper parameters were subsequently optimized to minimize the difference with the DFT data; the final copper parameters in this adaptation process are then: β_s +9.10 eV, β_p −42.93 eV and β_d −15.66 eV (CNDO) and β_s −27.57 eV, β_p −22.16 eV and β_d +6.27 eV (INDO). The excitation energies when using these parameters for the reduced states of the three proteins are given also in Tables 8.1.2. In all three cases, the HOMO now consists mainly of contributions from copper (~80 %) with approximately 15 % contributions from sulphur. The average absolute deviations for the ten lowest energies are then 0.10 (CNDO) and 0.15 (INDO) eV for the energy, and 0.008 (CNDO) and 0.006 (INDO) for the oscillator strengths.

Conclusions

The excitation energies of the active site of three copper proteins in the reduced state have been computed using Density Functional Theory, using the GRAC exchange-correlation potential. The ionization potentials of the active site, which are needed for this potential, were estimated from calculations on the active site using the SAOP xc-potential, which is known to give reliable results for them.

The obtained excitation energies were used to check the performance of the semi-empirical CNDO and INDO approaches for the computation of the excitation energies, and the copper atomic parameters needed for this approaches were checked. Two sets had been previously been used on small copper complexes, which give either too high or too low excitation energies. Adapting the copper parameters improves the computed excitation energies and oscillator strengths considerably, resulting in average absolute deviations from the DFT results of 0.10/0.15 eV for the energy and 0.006/0.008 for the oscillator strength.

8.2 *Oxidized states*

Excitation energies of copper proteins in the oxidized state

In the previous section, the copper parameters needed in the semi-empirical CNDO/INDO methods have been calibrated for a description of the reduced state of the active sites of three copper proteins. The lowest ten excitation energies computed with these parameters showed a small difference with the corresponding reference data, which had been obtained with Density Functional Theory (DFT)[1] calculations. As was already explained in the previous section, the excitation energies of the *oxidized* state can not be obtained directly with DFT; therefore the semi-empirical (CNDO/INDO) approach is used to compute them, using a Restricted Open Shell Hartree Fock (ROHF) wavefunction followed by Configuration Interaction (CIS). This method has been shown to give good results for organic molecules.[309]

TABLE 8.2.1. EXCITATION ENERGIES (EV) AND OSCILLATOR STRENGTHS USING "OPTIMIZED" (CNDO/INDO) PARAMETERS

m121h		*m121q*		*wildtype*	
CNDO[a]		*CNDO*[a]		*CNDO*[a]	
energy	*osc.str.*	*energy*	*osc.str.*	*energy*	*osc.str.*
0.35	0.0000	0.39	0.0000	0.31	0.0000
0.70	0.0000	0.51	0.0000	0.39	0.0000
0.97	0.0002	0.71	0.0000	0.84	0.0000
1.25	0.0001	0.88	0.0002	1.25	0.0004
3.01	0.0000	3.00	0.0000	3.07	0.0000
3.06	0.0000	3.05	0.0000	3.07	0.0000
3.10	0.0000	3.69	0.0476	3.24	0.0414
3.54	0.0403	3.86	0.0000	3.85	0.0000
3.88	0.0000	3.87	0.0000	3.90	0.0000
3.90	0.0000	4.65	0.3521	4.47	0.0002
INDO[b]		*INDO*[b]		*INDO*[b]	
energy	*osc.str.*	*energy*	*osc.str.*	*energy*	*osc.str.*
0.08	0.0000	0.09	0.0000	0.09	0.0000
0.11	0.0000	0.11	0.0000	0.09	0.0000
0.15	0.0000	0.13	0.0000	0.12	0.0000
0.20	0.0000	0.16	0.0000	0.47	0.0000
2.91	0.0000	2.89	0.0000	2.97	0.0000
2.96	0.0000	2.95	0.0000	2.97	0.0000
3.00	0.0000	3.28	0.0073	3.04	0.0063
3.09	0.0061	3.67	0.0000	3.66	0.0000
3.71	0.0000	3.69	0.0000	3.71	0.0000
3.73	0.0000	4.70	0.0000	4.84	0.3562

a) using optimized CNDO copper parameters (see Section 8.1)
b) using optimized INDO copper parameters (see Section 8.1)

The UV/Vis spectra of the oxidized states of the three proteins are characterized by intense absorptions in the visible region:[23] for M121H-azurin at 440 nm (2.818 eV) and 590 nm (2.101 eV) (with relative intensities of 2.1 vs. 1.0), for M121Q-azurin at 450 nm (2.755 eV) and 610 nm (2.033 eV) (with relative intensities of 1.0 vs. 8.0) and for *wildtype* azurin at 625 nm (1.984 eV). These absorption bands give the proteins their characteristic intense green and blue colors.

Optimized copper parameters

With the optimized copper parameters from the previous section, needed for the semi-empirical CNDO/INDO methods, the excitation energies of the oxidized state of the active site of the three copper proteins (M121H-azurin, M121Q-azurin and *wildtype* azurin) have been computed; see Table 8.2.1 for the ten lowest energies with corresponding oscillator strengths. The CNDO results with the new copper parameters (see Section 8.1) show a first excitation energy with considerable oscillation strength at too high energy (3.54, 3.69 and 3.24 eV for M121H, M121Q and *wildtype* azurin respectively) or too low wavelength (energies correspond to wavelengths of 350, 336 and 382 nm). For all three proteins, the second excitation energy with considerable oscillation strength is found at 4.64-4.68 eV (265-267 nm).

The INDO results with the new copper parameters (see Section 8.1) show a strange behavior with three to four excitation energies at very low energy (< 0.2 eV), a first excitation energy with non-zero oscillator strength at too high energy (or too low wavelength, at 401, 378 and 408 nm respectively for M121H, M121Q and *wildtype* azurin) and extremely low oscillator strength. The second excitation energy with non-zero oscillator strength is found at 4.80-4.84 eV (256-258 nm).

The excitation energies have also been computed with the CNDO/INDO methods using the standard copper parameters (see Table 8.2.2). Although there is some better agreement with the experimental data, there remains room for improvement.

Standard copper parameters

The CNDO results with the standard copper parameters result in excitation energies that are either too low (for M121H by 0.4-0.7 eV and for *wildtype* azurin by 0.1 eV) or too high (for M121Q by 0.5-0.7 eV), if the excitation energies with the largest oscillator strengths are taken into account. The relative values of the oscillator strengths corresponding to these excitations are in reasonable agreement with the experimental data[23]; for M121H azurin, the two lowest excitation energies with non-zero oscillator strength, the ratio of these strengths is 1:2.5 (experimentally 1:2.1), for M121Q azurin the ratio is 3.3:1 (exp. 8:1). For *wildtype* azurin one excitation energy with a large oscillator strength is found, while a second one with almost negligible strenght is observed, which agrees well with the experimental data. Finally, it should be mentioned that for M121H-azurin, a third excitation energy with significant oscillator strength is observed in the visible wavelength region (400-800 nm or 1.55-3.10 eV) at 2.45 eV with an oscillator strength equal in size to the one of the lowest excitation energy.

TABLE 8.2.2. EXCITATION ENERGIES[a] (EV) AND OSCILLATOR STRENGTHS USING STANDARD COPPER PARAMETERS

m121h		*m121q*		*wildtype*	
CNDO		*CNDO*		*CNDO*	
energy	*osc.str.*	*energy*	*osc.str.*	*energy*	*osc.str.*
0.90	0.0007	0.78	0.0001	0.46	0.0011
1.78	0.0228	1.58	0.0027	1.86	0.1227
2.12	0.0565	2.06	0.0044	2.06	0.0128
2.45	0.0207	2.52	0.1146	2.11	0.0005
3.04	0.0000	3.01	0.0000	2.20	0.0002
3.09	0.0000	3.06	0.0000	2.87	0.0072
3.11	0.0000	3.45	0.0343	3.09	0.0000
3.40	0.0842	3.85	0.0000	3.10	0.0000
3.78	0.0003	3.87	0.0000	3.92	0.0000
3.90	0.0000	4.48	0.0083	3.97	0.0000
INDO		*INDO*		*INDO*	
energy	*osc.str.*	*energy*	*osc.str.*	*energy*	*osc.str.*
0.74	0.0016	0.71	0.0004	0.51	0.0012
1.52	0.0203	1.31	0.0017	1.67	0.1038
1.71	0.0031	1.52	0.0023	1.78	0.0391
1.86	0.0266	1.72	0.0026	1.86	0.0025
2.52	0.0773	2.48	0.1309	1.90	0.0007
2.93	0.0000	2.91	0.0000	2.78	0.0039
2.98	0.0000	2.95	0.0000	2.99	0.0000
3.01	0.0000	3.68	0.0000	3.00	0.0000
3.61	0.0020	3.70	0.0001	3.74	0.0000
3.72	0.0074	3.78	0.0270	3.79	0.0000

The INDO results with the standard copper parameters show a somewhat better agreement with experiment for M121H-azurin, with excitation energies that differ by some 0.2-0.3 eV. However, the ratio of the oscillator strengths is less well predicted at 1:2.9 (experimentally 1:2.1). The third excitation energy with significant oscillator strength that was found with the CNDO method, is not present in the INDO results. For M121Q-azurin, the two lowest excitation energies with significant oscillator strengths are found at too high energy by 0.5-1.0 eV; the ratio of the relative oscillator strength (4.8:1) is however improved in comparison with the CNDO results (3.3:1), but are still far off from the experimental data (8:1). For *wildtype* azurin, two excitation energies with significant oscillator strength are found, which are found at too low energy in comparison with experimental data, which show only one peak.

In all of these calculations, the excitation energies were calculated for the active site *in vacuo*, thereby ignoring the influence of the (protein) environment on these properties, which can have a significant effect on the energies, as has been shown for the solvent effect on the $n \rightarrow \pi^*$ transition of acetone[196]. In the next section, this influence is taken into account.

Conclusions

The excitation energies of the active site of three copper proteins in the oxidized state have been computed using semi-empirical (CNDO/INDO) methods. The influence of using the standard copper parameters or the parameters optimized for the reduced state of these proteins (see Section 8.1), has been checked. The optimized copper parameters result in excitation energies with non-zero oscillator strength that are too high in energy (CNDO), or give oscillator strengths that are too low to account for the experimentally observed bright colors.

The calculations with the standard copper parameters show better agreement with the experimental data; although the excitation energies from the CNDO method are too low or too high by 0.2-0.7 eV, the ratio of the corresponding oscillator strengths is in good agreement with experiment. The excitation energies from using the INDO method are for M121H-azurin in better agreement with experiment, but for M121Q and *wildtype* azurin the agreement is less. The ratio of the oscillator strengths is, in comparison with CNDO results, improved for M121Q azurin, but for M121H and *wildtype* azurin the agreement is less. For the latter, even two excitation energies with significant oscillator strength are observed, while experimentally only one peak is observed.

The influence of the protein environment on the energies and oscillator strengths is ignored in these calculations, but it may have a significant effect; in Section 8.3, it will be taken into account.

8.3 *Protein environment*

The influence of the protein and solvent environment on the excitation energies of the active site in the reduced and oxidized state

Until here in this chapter, the excitation energies were computed for the active site of copper proteins *in vacuo*, thereby ignoring the influence of the protein and solvent environment on the energies and oscillator strengths. However, it is known that the solvent effect on excitation energies can be significant, as was shown for the n→π* transition of acetone in different solvents.[196] Similar significant effects might be expected for the presence of the protein environment surrounding the active site. Therefore the excitation energies were computed in the presence of (part of) the protein surrounding the active site, where all the residues within a radius of 10 Å from the center of the active site have been taken into account. Unlike the previous sections, where the coordinates of the active sites of the (two or four) molecules in the asymmetric unit have been averaged to obtain one active site geometry, in this section each molecule in the asymmetric unit is used separately and the energies averaged afterwards.

The calculations were performed using the Direct Reaction Field approach[133-136,138-142], where the active site has been treated at the semi-empirical CNDO/INDO (QM) level, the protein environment at a polarizable classical (MM) level, with a self-consistent coupling between the QM and MM systems. That is, the polarization of the QM system by the MM system, of the MM system by the QM system, and of the MM as well as the QM system by themselves are properly taken into account (see also Sections 2.3 and 2.4).

Reduced state

The computed excitation energies from the CNDO method with the optimized copper parameters for the reduced state (see Section 8.1) are given in Table 8.3.1. The presence of the protein environment brings the excitation energies down by some 0.3-0.5 eV, in comparison with the *in vacuo* results. The oscillator strengths remain more or less the same, with for M121H and M121Q-azurin the largest value found for the fifth excitation energy. For *wildtype* azurin, the excitation energy which had the largest oscillator strength (the seventh), now has almost completely lost it, and only one excitation energy with significant oscillator strength is left (the tenth).

The highest occupied molecular orbital (HOMO) consists mainly of contributions from copper (~80-90 %) with smaller contributions from the sulphur atom of cysteine 112 (~10 %) and the nitrogen atom of either one of the two histidine residues (~2 %). In comparison with the *in vacuo* results, these contributions are the same.

The INDO results show a similar behavior for the largest part of the excitation energies, but unlike the CNDO energies that were all decreased in comparison to the *in vacuo* results of Section 8.1, some INDO energies remain virtually the same, while others (for instance the lowest excitation of *wildtype* azurin) are increased in comparison with the *in vacuo* results of Section 8.1.

TABLE 8.3.1. EXCITATION ENERGIES (EV) AND OSCILLATOR STRENGTHS FOR REDUCED STATE OF ACTIVE SITES IN PRESENCE OF PROTEIN ENVIRONMENT

m121h		*m121q*		*wildtype*	
CNDO[a]		*CNDO*[a]		*CNDO*[a]	
energy	*osc.str.*	*energy*	*osc.str.*	*energy*	*osc.str.*
1.25	0.0010	0.90	0.0002	1.51	0.0006
1.39	0.0010	1.18	0.0001	1.62	0.0004
1.46	0.0001	1.32	0.0004	1.67	0.0018
1.50	0.0025	1.40	0.0000	1.68	0.0006
1.62	0.0175	1.62	0.0482	1.73	0.0010
1.72	0.0018	1.93	0.0052	1.83	0.0035
1.77	0.0005	2.01	0.0024	1.86	0.0006
1.81	0.0011	2.05	0.0030	1.95	0.0013
1.84	0.0043	2.16	0.0106	2.00	0.0000
1.87	0.0018	2.18	0.0077	2.10	0.0174
INDO[b]		*INDO*[b]		*INDO*[b]	
energy	*osc.str.*	*energy*	*osc.str.*	*energy*	*osc.str.*
1.40	0.0001	1.04	0.0001	1.79	0.0001
1.44	0.0002	1.09	0.0001	1.83	0.0003
1.46	0.0000	1.11	0.0000	1.86	0.0001
1.47	0.0004	1.14	0.0000	1.86	0.0002
1.52	0.0061	1.30	0.0189	1.90	0.0062
1.96	0.0006	1.90	0.0002	2.57	0.0001
1.97	0.0002	1.92	0.0001	2.63	0.0014
1.99	0.0016	1.94	0.0015	2.64	0.0005
1.99	0.0008	1.97	0.0042	2.65	0.0005
2.00	0.0044	1.99	0.0048	2.68	0.0066

a) using optimized CNDO copper parameters (see Section 8.1)
b) using optimized INDO copper parameters (see Section 8.1)

The oscillator strengths corresponding to the excitations are almost all decreased by a significant amount; for instance the fifth excitation of M121H azurin had *in vacuo* a strength of 0.018, while it is now decreased by a factor of three to 0.006. For M121Q azurin, the excitation with the largest oscillator strength is now the fifth excitation with a strength of 0.019, while *in vacuo* the largest strength was 0.054, which is also a decrease by a factor of three. For *wildtype* azurin, the factor with which the largest oscillator strength decreases is even larger (0.042 *in vacuo*, 0.007 here).

The contributions to the HOMO are, just like found with the CNDO method, mainly coming from copper itself (~90 %), with smaller contributions from Cys112 sulphur (~5-10 %) and His46/His117 nitrogens (~1 %).

Oxidized state

The excitation energies of the active sites in the oxidized state in the presence of the protein are given in Table 8.3.2. In comparison with the *in vacuo* results, a better agreement with experimental data[23] is found.

The excitations with the largest oscillator strengths from the CNDO results are observed at 2.25 and 2.88 eV for M121H-azurin, at 1.99 and 2.91 eV for M121Q-azurin and at 2.05 eV for *wildtype* azurin, which agree very well with the experimental numbers[23] of 2.10/2.82 (M121H), 2.03/2.76 (M121Q) and 1.98 (*wildtype*). Compared with the *in vacuo* results, this means a shift in energy of some 0.2-0.6 eV, either downwards or upwards. The ratio of the oscillator strengths for these excitations is in reasonable agreement with experiment, with values of 1:1.3 (M121H, exp. 1:2.1) and 4.9:1 (M121Q, exp. 8:1).

The contributions to the singly occupied HOMO are again coming mainly from copper (~70 %), with smaller contributions from Cys112 sulphur (~20-25 %) and His46/His117 nitrogen atoms (~1-5 %).

TABLE 8.3.2. EXCITATION ENERGIES[a] (EV) AND OSCILLATOR STRENGTHS FOR OXIDIZED STATE OF ACTIVE SITES IN PRESENCE OF PROTEIN ENVIRONMENT

m121h		*m121q*		*wildtype*	
CNDO		*CNDO*		*CNDO*	
energy	*osc.str.*	*energy*	*osc.str.*	*energy*	*osc.str.*
0.45	0.0003	0.41	0.0000	0.51	0.0002
1.28	0.0049	1.12	0.0014	1.84	0.0004
1.46	0.0251	1.49	0.0043	2.05	0.0767
1.65	0.0122	1.91	0.0037	2.07	0.0001
2.25	0.0325	1.99	0.0757	2.23	0.0008
2.29	0.0001	2.51	0.0000	2.48	0.0003
2.34	0.0035	2.83	0.0000	2.54	0.0050
2.56	0.0000	2.91	0.0155	2.64	0.0011
2.86	0.0040	3.17	0.0000	2.88	0.0049
2.88	0.0431	3.22	0.0007	3.00	0.0006
INDO		*INDO*		*INDO*	
energy	*osc.str.*	*energy*	*osc.str.*	*energy*	*osc.str.*
0.36	0.0006	0.32	0.0002	0.38	0.0005
1.04	0.0065	0.99	0.0012	1.73	0.0030
1.13	0.0017	1.06	0.0011	1.91	0.0714
1.44	0.0197	1.30	0.0026	2.02	0.0000
2.08	0.0437	1.90	0.0896	2.12	0.0003
2.14	0.0007	2.30	0.0000	2.30	0.0077
2.38	0.0000	2.52	0.0000	2.37	0.0059
2.42	0.0003	2.86	0.0001	2.50	0.0007
2.65	0.0009	2.96	0.0014	2.62	0.0002
2.95	0.0420	3.04	0.0175	2.85	0.0002
exp.[23]		*exp.*[23]		*exp.*[23]	
2.10	1.0 (*rel.*)	2.03	8.0 (*rel.*)	1.98	0.05[71]
2.82	2.1 (*rel.*)	2.76	1.0 (*rel.*)		

a) using standard copper parameters

The INDO results are similar to the CNDO results, but show a slightly larger deviation from the experimental numbers[23] (see Table 8.3.2). The INDO excitations with the largest oscillator strengths are found at 2.08 and 2.95 eV (M121H-azurin), 1.90 and 3.04 eV (M121Q-

azurin) and at 1.91 eV (*wildtype* azurin), which agrees reasonably well with the experimental numbers of 2.10/2.82 (M121H), 2.03/2.76 (M121Q) and 1.98 (*wildtype*) eV. Just like was observed for the CNDO method, the presence of the protein induces a shift in the excitation energies of some 0.2-0.8 eV.

The singly occupied HOMO consists in the INDO case of contributions from mainly copper (~70 %), Cys112 sulphur (20-25 %) and His46/His117 nitrogen atoms (1-5 %).

Conclusions

The influence of the presence of the protein on the computed excitation energies and corresponding oscillator strengths is investigated using the Direct Reaction Field approach, where the active site is treated with semi-empirical (CNDO/INDO) methods and the surrounding protein and solvent environment with a polarizable classical force field.

The calculations show a negligible effect on the contributions of (copper or Cys112 sulphur) atomic orbitals to the highest occupied molecular orbital for both the reduced and oxidized state, while the excitation energies are shifted by 0.2-0.8 eV. The CNDO results are in good agreement with experimental data, both for the excitation energies as well as the corresponding oscillator strengths. The INDO results show a reasonable agreement with experiments.

chapter nine

QM/MM APPROACH

active site geometries of copper proteins

CONTENTS

SUMMARY

Using hybrid QM/MM calculations to optimize the geometry of the active site of metalloproteins in the presence of the protein and a layer of solvent molecules is presented in this chapter. A new link model for use in these QM/MM studies is presented that minimizes the influence of the introduction of artificial capping atoms is presented, which is used in subsequent optimizations of the active sites of wildtype, mutated and metal-substituted azurin.

9.1 *AddRemove*

A new link model for use in QM/MM studies

Hybrid QM/MM methods[149] split a system under study up in two parts: the electronically more important one that is treated with quantum mechanics (*real* QM system) while the remaining part is treated on a classical level (MM system). The division into the two subsystems is sometimes trivial as in the case of solvation of an organic compound where one puts the organic compound in the QM system and the solvent in the MM system. However, in biochemical systems the division often results in bonds that cross the QM/MM boundary. Several approaches have been used to handle this situation (see ref. 149), like for instance localized orbitals[310], pseudopotentials[150,311-313], but the easiest and most commonly chosen solution is to use link atoms[314,315]. In the latter concept one introduces a (artificial) capping atom in the quantum system (comprising then the *capped* QM system), which is usually taken to be a hydrogen atom. Singh and Kollman[314] (who originally proposed the concept) didn't put any constraints on the position of the capping atom, thereby adding three degrees of freedom to the system. To avoid the addition of these unwanted and artificial degrees of freedom, many later developed link models impose the capping atom LC, the quantum atom LQ and the classical atom LM to have a mutual coordinate dependence (for instance to lie on a straight line), thereby removing again the additional degrees of freedom (see Figure 9.1.1).

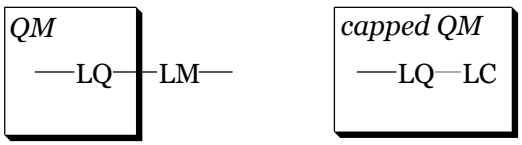


FIGURE 9.1.1. SCHEMATIC REPRESENTATION OF LINK BOND

Although there are many link models in the recent literature[151-154,316], the newly developed model as described in this section (AddRemove), will be compared primarily with the currently implemented model in the ADF[117] program, as this model is more or less representative for the other models. After introducing the new model, a comparison is made also with other (well known) models.

In the current implementation in ADF[117] (IMOMM/ADF)[155-157], the degrees of freedom of the real classical atom LM are removed: it "follows" the capping atom LC with its position depending on the positions of the LQ and LC atoms as:

$$\vec{r}_{LM} = \alpha\ \vec{r}_{LC} + (1-\alpha)\ \vec{r}_{LQ} \qquad (1)$$

In this equation, α is an arbitrary constant parameter but its value is difficult to generalize or choose. The energy depends on the value chosen for it and one can compare energies between different molecules only if the αs are chosen equal for both.

The capping atom interacts only within the *capped* QM system and has no interactions with the MM environment. Moreover, the real classical LM atom does not interact with the real QM system; its interactions with the QM system are "replaced" by the QM interactions of the capping atom LC. The real classical LM atom of course does have MM interactions with the rest of the MM system, but they

are not used for updating its position (since it is constrained to "follow" the LC atom).

Before doing the QM calculation, the MM system is optimized first. In this part, the atoms in the *real* QM system are kept fixed at their positions, and interact with atoms in the MM system through classical MM interactions. In the IMOMM/ADF implementation[155-157], also the classical LM atoms are kept fixed in this part. A schematic example for the optimization scheme is given on the left in Figure 9.1.2.

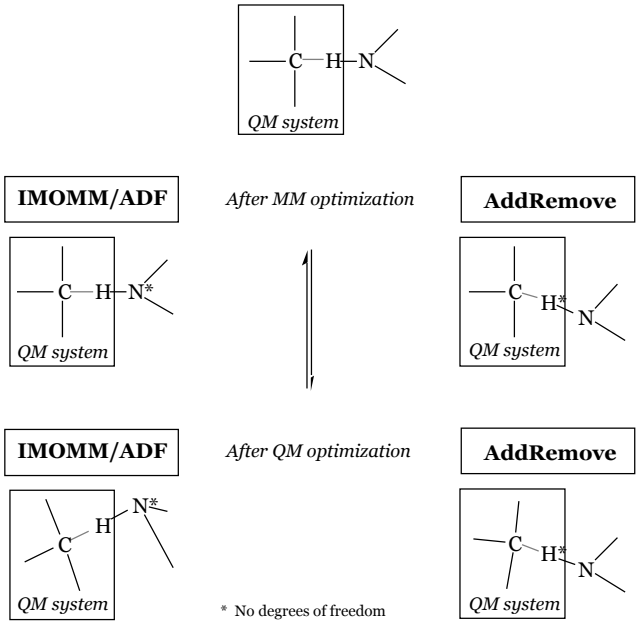


FIGURE 9.1.2. QM/MM OPTIMIZATION SCHEME

After the MM optimization has ended, the QM calculation is performed, and the QM energy, gradients and charges calculated. Then the MM system is optimized again before starting a new QM cycle, and so on, until the QM system is fully optimized.

One of the advantages of the IMOMM and AddRemove model is that the interactions between the QM and MM system are treated at the same level of theory as the interactions within the MM system itself. Then, if one uses for those interactions the AMBER95 or GROMOS96 force fields[126,127] (which is constructed for treating the interactions within proteins, and is known to perform well for them), one is certain that the interactions (in the MM system itself as well as between the QM and MM systems) are treated properly.

AddRemove methodology

In the AddRemove model, the position dependence is reversed: it is the capping atom LC that "follows" the classical atom LM, and not vice versa:

$$\vec{r}_{LC} = \vec{r}_{LQ} + R_{eq}\vec{u}_{LQ-LM} \tag{2}$$

Here, $\vec{u}$ is a unit vector pointing from atom LQ to atom LM (see Figure 9.1.1 for atom labels), and $\boldsymbol{R_{eq}}$ is the equilibrium distance for the LQ-LC bond it would have in the classical force field that is used for

the MM system. This has the major advantage that one does not remove the degrees of freedom of the real classical atom LM (as is the case in the IMOMM/ADF model[155-157]). Instead, the artificial degrees of freedom of the capping atom LC are removed. Also, unlike the IMOMM/ADF implementation[155-157], the energy does not depend anymore on some arbitrary parameter α. Furthermore, the real classical LM atoms interact normally with the *real* QM system, and are allowed to move in the optimization of the MM system (see also the right hand side of Figure 9.1.2). Then, before the QM calculation starts, the positions of the capping atoms are updated with eq. (2) from the current positions of the LM atoms. After the QM energy and gradients have been obtained, they are corrected for by removing the interactions of the capping atoms: the MM interactions, which the capping atoms would have with the atoms in the real QM system if it were a classical atom, are subtracted. That is, the capping atoms are Added and Removed. Schematically, this would result in the following energy expression for the interaction between the QM and MM systems:

$$U^{QM/MM} = \sum_{\substack{j\ MM \\ i\ realQM}} \sum_{j \neq LM} U^{MM}(i,j) + \sum_{n=1}^{Nlinks} \sum_{i\ realQM} U^{QM}(i,LC_n) - U^{MM}(i,LC_n) + U^{MM}(i,LM_n) \qquad (3)$$

The first summation runs over all classical atoms that are not involved in the link bond, which couple normally with the *real* QM system through standard MM interactions. The second part is specific for the atoms involved in the link bonds. Here, the "quantum" interactions of the capping atoms LC with the *real* QM system ($U^{QM}(i,LC_n)$) are corrected for by subtracting its MM counterparts ($U^{MM}(i,LC_n)$), and replaced by the MM interactions of the classical LM atom with the *real* QM system ($U^{MM}(i,LM_n)$). Of course, it is not possible to project out the $U^{QM}(i,LC_n)$ interaction, but schematically this is justified. Moreover, after summation and combination with the "QM energy" of the *real* QM atoms, simply the total QM energy (of the capped QM system) is recovered.

For the forces working on the atoms, the same procedure is followed. That is, the following expression is valid for the gradients of the interaction energy:

$$g^{QM/MM} = \sum_{\substack{j\ MM \\ i\ realQM}} \sum_{j \neq LM} g^{MM}(i,j) + \sum_{n=1}^{Nlinks} \sum_{i\ realQM} g^{QM}(i,LC_n) - g^{MM}(i,LC_n) + g^{MM}(i,LM_n) \qquad (4)$$

This expression follows directly from using eq. 3 for the energy expression. Note that this expression is one of the things where the AddRemove model differs from other link models (like ONIOM[151-153]); this topic is discussed in more detail in the **Comparison** part.

For convenience, the gradient of the artificial LC atoms is set to zero, since any resulting nonzero gradient would be the result of the inability of the MM force field to represent QM interactions; this effect is ignored also for the MM system. Moreover, the position of the capping atom is not a degree of freedom, and leaving the gradients of the capping atoms nonzero would affect only the convergence of the optimization procedure of the QM system, **not** the positions of the atoms.

The electrostatic interactions between the *real* QM and MM systems are obtained in a classical manner, but can be handled in an iterative way (*updated electrostatic coupling*): during the optimization procedure, the charges of the QM system are updated after each QM optimization cycle and used in the next MM optimization cycle. The charges for the QM system are obtained from a recently developed charge analysis, which is derived from an atomic multipole expansion (the Multipole Derived Charge analysis)[183]. In the ADF program[117], the Coulomb potential can be obtained from a set of atomic multipoles[118] (together with a short range function). These atomic multipoles are

reproduced *exactly* by the charges, and therefore also the molecular multipoles as well as the Coulomb potential. In this approach, we take as much data from the QM calculations as possible.

Comparison with other link models

The expression for getting the interaction energy seems similar to ones used by other link models like for instance the ONIOM model[151-153], when the latter is used with 2 layers. However, this similarity is only present in the energy expression. The specific implementation of this expression differs to a great extent.

Unlike the ONIOM model[151-153], the degrees of freedom of all real atoms are completely free, i.e. all *real* atoms are completely free to move. This is in contrast to the ONIOM method, where not only the bond between the real quantum atom LQ and the capping atom LC, but also the bond between the *real* atoms LQ and LM is frozen at a certain predefined distance. This is similar to the introduction of the IMOMM/ADF[155-157] model α-parameter, upon which the energy and geometry depends. In fact, it is found that depending on the specific value given for this parameter (g parameter in ONIOM terminology), the bond between the *real* LQ and LC atoms ranges from 1.46 to 1.67 Å for a CC bond when the parameter α is varied from 0.5 to 1.0[151].

More importantly, the gradient expressions don't follow naturally from the energy expressions; unlike our implementation, where not only the energy but also the gradient is corrected (see eqs. 3 and 4), in the ONIOM method the gradient of the capping atoms is *projected* onto the LQ and LC atoms. This not only introduces the coordinates of the capping atoms as degrees of freedom, but more importantly, introduces inconsistencies in the total picture of the system. Finally, unlike the ONIOM method, which needs several calculations for several systems on a different level that are coupled in a not always well-defined manner, the AddRemove model can be used directly within a QM scheme. The only modifications needed are the optimizations of the MM system before every QM step, followed by adding the QM/MM interaction energies and forces to the QM and MM forces to get the total energies and forces working on the atoms. In this respect, the AddRemove model is more similar to the QMPot model by Sauer et al.[154]. However, unlike their work, the interactions between the QM and MM systems are simply taken from a classical force field (AMBER95[126]) that is known to perform well for treating interactions within proteins; therefore, no additional assumptions have to be made.

In the **Results** we will compare results from using the AddRemove model with those found for the IMOMM/ADF model[155-157]; although there are differences between the IMOMM[155-157] and ONIOM[151-153] energy expressions, both use the same gradient expressions (by projecting the gradient of the artificial link atoms onto its neighboring atoms). As the geometry obtained in a geometry optimization depends only on the gradient, *not* on the energy, the IMOMM results are representative also for a 2-layered ONIOM model.

Results

In Sections 9.2-9.4, the results of using a QM/MM approach for getting the active site geometries of copper proteins are reported, where the AddRemove model has been used to couple the QM and MM systems. In those sections, the QM system consists of the complete active site that is surrounded by a MM system, which consists of the rest of the protein as well as a layer of solvent molecules surrounding the protein.

In order to check the validity of the AddRemove model for treating the interactions involved in cutting such a biochemical system into a QM and MM system, a few amino acid residues were treated either by a full QM, or a QM/MM description. As the main objective in the study on copper proteins (see Sections 9.2 to 9.4) is the active site geometry, in this section only the geometry of these residues will be discussed. Moreover, as it is the gradient expression (see equation 4) in which the AddRemove differs from other link models (like ONIOM[151-153]), and the geometry is determined completely by the gradient, not the energy, this enables a meaningful comparison between AddRemove and other models.

For the QM/MM description, both the IMOMM/ADF[155-157] and the AddRemove model have been used. The residues studied are methionine (Met), cysteine (Cys) and tyrosine (Tyr). The amino acid residues were described with a dipeptide surroundings, where on both sides of the backbone connection, the peptide bond is included explicitly. The residues were studied with either the complete molecule in the QM system (*full QM*), or with the QM system cut off at the C-alpha position with hydrogens added to complete the link bonds (*QM/MM)*. In the QM/MM description, the electrostatic coupling of the QM and MM systems was either done by a *simple* or an *update* approach. In the latter approach, the charges of the QM atoms are updated after every QM geometry optimization cycle, while in the *simple* approach the standard AMBER95 charges were used.[126]

The calculations were performed with the ADF program[117,187], using the Becke[120]-Perdew[121] exchange-correlation potential in a triple zeta valence plus polarization (TZP) basis set. In another study[317], where a number of exchange-correlation potentials available in ADF were tested in several standard basis sets, it was found that the Becke-Perdew xc-potential in that basis set predicts the bond distances of a certain set of molecules with an average absolute deviation of 1.24 pm. This value gives an indication of the accuracy of the bond distances that could be reached.

TABLE 9.1.1. LINK BOND DISTANCES (PM) FROM *FULL QM* AND *QM/MM* STUDIES

	full QM	*IMOMM simple el.st.*	*IMOMM updated el.st.*	*AddRemove simple el.st.*	*AddRemove updated el.st.*
cysteine					
Cα-N	146.0	151.7	151.1	147.1	146.7
Cα-C	155.6	157.8	158.9	154.3	155.4
Cα-Cβ	153.3	153.6	153.5	152.5	152.6
methionine					
Cα-N	145.8	152.1	151.6	147.3	145.7
Cα-C	154.2	157.1	157.7	153.9	155.2
Cα-Cβ	154.7	154.3	154.4	153.4	153.2
tyrosine					
Cα-N	146.9	151.5	151.8	147.0	147.8
Cα-C	155.9	157.7	157.6	154.4	153.4
Cα-Cβ	153.1	154.5	154.7	152.7	153.0

The most relevant property to look at in these calculations, is the distance between the *real* atoms involved in the link bonds; as the system is cut off at the C-alpha position, there are two distances per

molecule, which result from the bonds of the C-alpha to the peptide-bond on either side (C_α-$N_{backbone}$ and C_α-$C_{backbone}$ distances). For convenience, the distance between the C_α and C_β atoms are also given to be able to check the performance of the methods in that respect. The distances for the *full QM* and IMOMM/AddRemove *QM/MM* descriptions are given in Table 9.1.1. These results show a rather poor performance for the IMOMM/ADF method[155-157] to describe the link bonds; the average absolute deviation in the C_α-$C_{backbone}$ bond is 2.3 pm (*simple* electrostatic coupling) or 2.8 pm (*updated* electrostatic coupling), while it is even higher for the C_α-$N_{backbone}$ bond: 5.5 (*simple*) and 5.3 (*updated*) pm. The other two *regular* bonds of the C-alpha atoms are well described with an average absolute deviation of 0.7 pm (C_α-C_β) and 0.1 pm (C_α-H; not shown here in the Table). Taking the C_α-C_β bond as a reference for the accuracy that could be reached by putting part of the molecule in the MM system, this means an increase by a factor of 3.1-4.0 for the C_α-$C_{backbone}$ bond deviation and 7.1-7.3 for the C_α-$N_{backbone}$ bond.

The AddRemove model performs much better in that respect with average absolute deviations of 0.9 (*simple*) or 0.6 (*updated*) pm for the C_α-$N_{backbone}$ bond, 1.0 (*simple*) and 1.2 (*updated*) pm for the C_α-$C_{backbone}$ bond. The regular bonds are equally well described as indicated by the average absolute deviation of 0.8 pm for the C_α-C_β bond, and < 0.1 pm for the C_α-H bond (data not shown in Table). By comparing again the accuracies of the link bonds with the one found for the C_α-C_β bond deviation, we find factors of 1.3-1.6 for the C_α-$C_{backbone}$ bond and 0.8-1.2 for the C_α-$N_{backbone}$ bond, which is a substantial improvement over the results of the IMOMM model[155-157].

Conclusions

The division of a system under study in a quantum mechanical and a classical system in QM/MM calculations is sometimes very natural (as in the case of solvation of an organic compound, where the organic compound can be treated at the QM level and the solvent at the MM level), but a problem arises in the case of bonds crossing the QM/MM boundary. A new link model is presented that uses a capping (link) atom to satisfy the valences of the quantum chemical system, with the position of the *artificial* capping atom depending on the positions of the real atoms involved in the link bond. Therefore, the introduction of the capping atom does not lead to additional degrees of freedom; moreover, the degrees of freedom of all *real* atoms are kept (unlike other link models). Furthermore, the introduction of this artificial atom is corrected for after every QM optimization cycle by subtracting the classical molecular mechanics interactions with the real QM system it would have if it were a classical atom. Charges of the *real* QM atoms can be updated after every QM geometry cycle, and used in the electrostatic coupling between the QM/MM systems in the subsequent optimization of the MM system.

The new model is compared to other (well-known) link models, and tested in comparison with the other link model implemented in the ADF program (IMOMM/ADF) on a few amino acid residues. In these test calculations the residue is described either by a complete QM description (*full QM*) or by a QM/MM description, where the QM system of the residue has been cut off at the C-alpha position. For the QM/MM description, both the AddRemove and IMOMM model has been used, either with *simple* or *updated* electrostatic coupling between the QM and MM systems. The distances between the C-alpha and the other two real atoms involved in the link bonds, differ only marginally when using either one of these two electrostatic couplings. This is a positive result, especially in view of the systems where the QM/MM method will be applied to (metalloproteins). For those systems, there are no predefined charges in the force field, and one has no other option than to use the updated ones.

A large difference is found between the performance of the IMOMM/ADF and AddRemove models. Taking the deviation for predicting the distance of the regular C_α-C_β bond (0.7 pm) as a reference value when part of the system has been put in the MM system, the IMOMM model gives a rather poor performance: the deviation increases by a factor of 3.1-4.0 for the C_α-$C_{backbone}$ bond and 7.1-7.3 for the C_α-$N_{backbone}$ bond. The AddRemove model gives a much better description; the deviation changes only slightly, by a factor of 1.3-1.6 for the C_α-$C_{backbone}$ bond and 0.8-1.2 for the C_α-$N_{backbone}$ bond.

9.2 *Wildtype azurin*

Optimization of the active site geometry in both the reduced and oxidized state

There have been many discussions over the last decades about the influence of the protein environment on the active site of copper proteins.[55,108,109] It has been argued that the protein forces the active site in a strained conformation, which is a compromise between the favored geometry of the reduced and oxidized state. Recently, this view has been challenged by Ryde et al.[11-14,17,20], who performed extensive studies on small model systems of the active site of type-1 copper proteins. They found active site geometries that are remarkably similar to the ones observed in plastocyanin, a typical type-1 copper protein that lacks the axial Gly residue observed in azurin. However, one feature that could not be reproduced by these model systems is the distance of the axial Met ligand, which was nearly always observed at too close a distance from copper. In another investigation by de Kerpel et al.,[104] the same procedure has been applied to study *wildtype* and Co-substituted azurin. Again problems were encountered for the axial ligands of the *wildtype* azurin model, with a distance of 7.18 Å for Met121 and 2.34 Å for Gly45. The Met121 residue has moved away completely and formed a hydrogen bond to the H_ε-atom of one of the imidazole groups.[104]

TABLE 9.2.1. OPTIMIZED DISTANCES (Å) OF *WILDTYPE* AZURIN ACTIVE SITE *IN VACUO I*

	Starting structure	*X-ray*[280] *reduced state*	*in vacuo I reduced state*	*X-ray*[7] *oxidized state*	*in vacuo I oxidized state*
Cu-Gly45[a]	2.955	3.021 ± 0.077	3.196	2.967 ± 0.093	2.981
Cu-His46[a]	2.064	2.133 ± 0.093	2.138	2.076 ± 0.060	2.038
Cu-Cys112[a]	2.267	2.293 ± 0.013	2.229	2.237 ± 0.044	2.177
Cu-His117[a]	1.978	2.095 ± 0.091	2.066	2.011 ± 0.069	2.012
Cu-Met121[a]	3.164	3.247 ± 0.069	3.459	3.149 ± 0.070	3.214
Cu-N2S[b]	0.043	0.092 ± 0.013	0.098	0.083 ± 0.051	0.100

a) distance of Cu to closest atom in this residue;
b) distance of copper from plane made by the S_γ(Cys112), N_δ(His46) and N_δ(His117) atoms

In Section 6.3, it was shown that with the Becke[120]-Perdew[121] xc-potential in a larger basis set, the potential energy surface of the axial copper-Met121 distance is very flat in the range from 3.0 to 4.0 Å. Therefore, just like observed by de Kerpel et al.[104], the Met121 residue has a large freedom to move and no well-defined energy minimum is observed. In contrast, for the other axial group, Gly45, a well defined energy minimum was observed around 2.9 Å. However, the Gly45 residue is connected to one of the in-plane ligands (His46) and as only the Gly45 residue was allowed to move in Section 6.3, the localization of such an energy minimum may have been forced by the connection to His46. A geometry optimization of the whole active site should give an indication of whether the Gly45 energy minimum is forced by the His46 connection, or not. Moreover, it can provide some more insight on the role of the protein on the active site geometry. Therefore, starting from the structure that was used in Section 6.3, a full geometry optimization was performed without any constraints on the atoms. The optimized copper-ligand distances are given in Table 9.2.1. Given in Table 9.2.1 are also the statistical errors from

the X-ray data[7,280] obtained after averaging the distances over the four molecules in the asymmetric unit. A remarkable similarity in these numbers is observed, apart from the distances of the axial groups in the reduced state. The difference with the X-ray data is at the edge of the uncertainty in the X-ray data, which is estimated to be 0.1-0.2 Å. Still, a satisfactory agreement between experimental and calculated distances is obtained. The difference between the two Histidine distances is properly predicted and the decrease in distance in going from the reduced to the oxidized state is observed as well.

Performing the geometry optimization with another starting structure for the active site (also taken from X-ray data[a 5]) however can have a marked influence on the obtained geometry (see Table 9.2.2), which indicates that for the active site *in vacuo* local minima exist with virtually the same energy.

TABLE 9.2.2. OPTIMIZED DISTANCES (Å) OF *WILDTYPE* AZURIN ACTIVE SITE *IN VACUO II*

	Starting structure	*in vacuo II reduced state*	*in vacuo II oxidized state*
Cu-Gly45[a]	3.066	3.007	2.480
Cu-His46[a]	2.208	2.050	2.040
Cu-Cys112[a]	2.216	2.217	2.180
Cu-His117[a]	2.362	2.126	2.023
Cu-Met121[a]	3.007	4.738	4.317
Cu-N2S[a]	0.056	-0.053	-0.130

a) see subscript Table 9.2.1 for definition

In this case, the Met121 residue has moved away from the copper completely, and is found at 4.7 Å in the reduced state and at 4.3 Å in the oxidized state. The flat potential energy surface between 3.0 and 4.0 Å for the copper-Met121 distance may be a large contributor to this drifting away of Met121. At the same time, Gly45 moves closer to the copper, in the oxidized state it is even found at 2.5 Å. The differences between calculated and experimental copper-residue distances are now too large to be "accounted for" by the experimental uncertainty. Therefore, it has to be concluded that if one wants to obtain good active site geometries, one can not ignore the influence of the protein: it should be included in the optimizations.

Hybrid QM/MM calculations

In principle, if one wants to include the whole protein in the optimizations, one could treat it completely by quantum chemical methods. However, even with the linear scaling techniques[118] that are currently in use and despite the enormous amount of computing power, performing a geometry optimization on a system of some 2000 atoms in an acceptable basis set is probably still awkward to do. Moreover, precious CPU time would be wasted as a detailed and therefore expensive description is used for large regions in space, which could be very well described by classical interactions by using specially designed protein force fields like AMBER95[126] or GROMOS96[127]. Therefore, a more appropriate and efficient way to include the protein in the geometry optimizations is by performing

[a] One of the molecules in the asymmetric unit of F114A-azurin, which shows a large similarity with *wildtype* azurin, apart from the elongated Cu-His distances

hybrid QM/MM calculations (see Section 2.3). In this case, one puts the place of interest (active site) in the QM region while the rest of the protein and solvent molecules are put in the MM system. In Section 9.1, a new link model (AddRemove[158]) to connect the two systems has been described, which uses capping atoms to satisfy the valency of the QM system. These capping atoms are necessary in the case of proteins, when covalent bonds are crossing the QM/MM boundary. The introduction of these capping atoms is artificial, and if not accounted for, would lead to an artificial increase of the degrees of freedom, as well as QM interactions that should not be present. Therefore, after every QM cycle, the QM interactions of these artificial capping atoms with the real QM atoms are "removed" by subtracting their counteracting MM interactions (see Section 9.1). The positions of the capping atoms are constrained to lie in the direction of the covalent bond crossing the QM/MM boundary, and therefore no artificial degrees of freedom are added to the system. Therefore, the capping atoms are "added" and "removed".

A hybrid QM/MM calculation has been performed using the QM/MM code[155] in the ADF program[117,187], which has been adapted and optimized in order to be able to treat proteins[a]. A conjugate gradient optimization scheme[1] has been added and used, the calculation of the forces has been parallellized and the memory allocation scheme of the van der Waals interactions adjusted to make it use the available memory more efficiently. The Becke[120]-Perdew[121] exchange-correlation potential has been used in a TZP basis set, with an integration accuracy of 6.0. For the construction of the ADF[117,187] inputfile, the pdb2adf[318] program has been used. After inclusion of the protein, the total number of QM atoms was 61, the total number of protein atoms 1930 and total size of the system ranging from 2200 to 14000, depending on whether a shell of water molecules surrounding the whole protein has been added. In order to keep the molecules from evaporating from the system, a (spherical) wall force has been introduced that acts to keep the atoms in an imaginary spherical box with a radius of 33 Å. The AddRemove[158] model (Section 9.1) has been used to couple the QM and MM system, with "updated" electrostatic coupling. This means that after a QM cycle, the charges on the QM atoms have been updated to their MDC-q values, which come from the Multipole Derived Charge analysis[183]. These charges are subsequently used in a classical electrostatic fashion to couple the QM and MM system.

The inclusion of the protein has a positive effect on the optimized copper-residue distances (see Table 9.2.3). Not only is again a good agreement observed between calculated and experimental[7] copper-residue distances, the same active site geometry is observed when started from another starting structure. Apart from the axial Met121, all residues are found closer to copper than in the X-ray data[7,280], and agree well with EXAFS data[319] that put the in-plane ligands closer to copper than X-ray does. The axial groups are too far away from copper to be "seen" by EXAFS[319].

The decrease in these distances in going from the reduced to the oxidized state is well predicted by the QM/MM calculations. Unlike the X-ray data[7,280], which place the two in-plane Histidines at distances that differ by some 0.04-0.06 Å, the QM/MM calculations predict that they are at much more similar distances from copper, with differences of 0.02 Å or less. Cys112 is positioned at 2.24 Å from copper in the reduced state and 2.17 Å in the oxidized state, which is at still somewhat larger distances than in other type-1 copper proteins like plastocyanin, but compares very well to EXAFS studies that place it at around 2.15 Å in the oxidized state and at 2.20 Å in the reduced state[25,319]. The decrease in this distance by about 0.05 in going from the reduced to the oxidized state is a feature preserved in the EXAFS and X-ray data, as well as in the QM/MM calculations.

[a] Adaptation and optimization of QMMM code by M. Swart for this thesis

TABLE 9.2.3. QM/MM OPTIMIZED DISTANCES (Å) OF *WILDTYPE* AZURIN ACTIVE SITE

	X-ray[280] *reduced state*	*QM/MM reduced state*	*X-ray*[7] *oxidized state*	*QM/MM oxidized state*
Cu-Gly45[a]	3.021 ± 0.077	3.003	2.967 ± 0.093	2.855
Cu-His46[a]	2.133 ± 0.093	2.069	2.076 ± 0.060	2.017
Cu-Cys112[a]	2.293 ± 0.013	2.238	2.237 ± 0.044	2.169
Cu-His117[a]	2.095 ± 0.091	2.076	2.011 ± 0.069	1.995
Cu-Met121[a]	3.247 ± 0.069	3.393	3.149 ± 0.070	3.178
Cu-N2S[a]	0.092 ± 0.013	0.065	0.083 ± 0.051	0.094

a) see subscript Table 9.2.1 for definition

Also for the angles a good agreement is observed between calculated and experimental[7,280] (X-ray) data (see Table 9.2.4). The difference between the two is at maximum some 6 degrees, which is well within the limits posed on the angles by the experimental uncertainty of the X-ray coordinates. Again a good agreement between calculated and experimental results is observed when looking at the difference in angles in going from the reduced to the oxidized state. In nearly all cases when an angle is found by X-ray data to decrease (or increase) upon oxidation, the same trend is observed by the QM/MM calculations.

TABLE 9.2.4. QM/MM OPTIMIZED ANGLES (°) OF *WILDTYPE* AZURIN ACTIVE SITE

	X-ray reduced state	*QM/MM reduced state*	*X-ray oxidized state*	*QM/MM oxidized state*
S_{112}-Cu-N_{46}	133	137	132	130
S_{112}-Cu-N_{117}	122	123	122	127
N_{46}-Cu-N_{117}	105	99	105	102
S_{112}-Cu-S_{121}	113	115	110	111
S_{112}-Cu-O_{45}	99	95	99	94
S_{121}-Cu-O_{45}	145	149	148	151
N_{117}-Cu-O_{45}	87	84	89	85
C_{45}-O_{45}-Cu	128	125	130	130
C_{112}-S_{112}-Cu	106	104	109	110
C_{121}-S_{121}-Cu	137	128	140	131
C^{γ}_{46}-N_{46}-Cu	130	129	131	129
C^{ε}_{46}-N_{46}-Cu	126	125	125	124
C^{γ}_{117}-N_{117}-Cu	128	125	130	125
C^{ε}_{117}-N_{117}-Cu	127	129	124	128

Despite their extensive work on model systems[11-14,20,103,104], also Ryde et al. came to the conclusion that in order to get good active site geometries, the protein has to be included by using QM/MM calculations (for which they use the ComQum label)[17]. Especially if one wants to obtain the axial groups at the correct position, this seems to be of vital importance. Although their ComQum results perform better than their model studies in this respect for oxidized plastocyanin, it still does

not place the axial methionine at the right position in the reduced state. However, was the distance underestimated in the model system by some 0.5 Å, in the ComQum calculation it is overestimated by almost the same amount. The same problems with the axial ligands is observed in their study on azurin, where the copper-Met121 distance is too short by 0.3 Å in the reduced state and too long in the oxidized state by 0.5 Å. Also the decrease in copper-Cys112 distance by about 0.05 Å upon oxidation is less pronounced in their ComQum calculations. As they also use the AMBER95[126] force field for describing the interactions in the MM system, the difference with our results can only be coming from either the use of the B3LYP[122] xc-potential or the smaller basis sets.

Spin density distribution

Atomic spin densities can be obtained by taking the difference between the atomic charges resulting from either the spin-up or spin-down density. Like mentioned in Section 3.1, the Multipole Derived Charges[183] are obtained from both densities, and therefore in a spin-unrestricted calculation provide not only the atomic charge but also the atomic spin density.

The spin density present in the oxidized state is delocalized in the active site, with large amounts present on copper (20.7 %), the sulphur of Cys112 (52.7 %), the sulphur of Met121 (9.3 %) and the nitrogens of His46 (4.8 %) and His117 (5.8 %), while on the oxygen of Gly45 only a tiny amount is observed (0.8 %). These results therefore indicate that although the Met121 is at a long distance from the copper atom, it is weakly covalently bonded to it, while Gly45, which is located closer to copper than Met121, is not covalently bonded or, at best, only very weakly. Also some tiny amounts of spin density are found on nearby protons, as in the case of Met121-Hε (~0.5 %) and Cys112-Hβ (~0.8 %), His46-Hα/β (0.3-0.4 %), His117-Hβ (~0.5 %) and His117/46-Hδ/ε (~0.1 %).

Hydrogen bonds

One of the more important residues in the nearby surroundings of the active site seems to be the Asn47 residue, which is preserved at this position in all azurin molecules and other type-1 copper proteins found so far. It makes five hydrogen bonds, two of which are to residues not directly related to the copper site, two to Ser/Thr113 with its sidechain oxygen and nitrogen (at 2.95/2.87 Å) and one to the Cys112 sulphur with its backbone amide nitrogen at a N-S distance of 3.44 Å. These are all conserved in the QM/MM optimizations, as is another hydrogen bond to the Cys112 sulphur, by the backbone nitrogen of Phe114 with a N-S distance of 3.79 Å. Also, hydrogen bonds are observed connecting the backbone nitrogen of Cys112 and the backbone oxygen of Met121 at 2.96 Å, the backbone nitrogen of Met121 and the backbone oxygen of His117 at 2.75 Å, and the backbone nitrogen of His117 and the backbone oxygen of Phe114 at 3.03 Å.

Conclusions

The optimization of the geometry of an active site *in vacuo* does not always lead to structures corresponding to the active site as it is observed in the protein. The influence of the protein on the geometry is found to be a factor that can not be simply ignored. This is especially valid for the axial groups in type-1 copper proteins, especially Met121 that has an enhanced freedom to move *in vacuo* (see also Section 6.3). Without the inclusion of the protein, the active site geometry may sometimes be reasonably represented, but in many cases the site deforms. After inclusion of the protein in the geometry optimization by hybrid QM/MM calculations, a very good agreement is observed between

calculated and experimental copper-residue distances and angles. In these QM/MM calculations, the system of interest (active site) is put in the QM system while the rest of the protein and solvent molecules is placed in the MM system. The results presented here give a better description than the ComQum results by Ryde et al.

9.3 *Metal substitution and mutants*

The effect of metal substitution and mutations on the active site geometry

In the previous section, a hybrid QM/MM calculation was used to optimize the active site geometry of *wildtype* azurin. However, many experimental studies have focused not only on *wildtype* azurin, but also on metal-substituted azurin[4,33,36,54,59,80,104,320-322] or mutants of it[3-5,23,27,30,31,35-38,42,48,61,63,75,77,80-84,88,319,322-326]. The metal-substitution is either found by accident, as in the case of zinc-azurin[59] that has been found as a colorless by-product in the heterologous expression of *wildtype* azurin, or made on purpose. In the latter case, copper is substituted for either zinc, cobalt, nickel, manganese, cadmium or mercury, for a structural or spectroscopic characterization by different techniques like paramagnetic NMR spectroscopy, X-ray diffraction or UV-VIS absorption spectroscopy. Another widely used technique to study the function of amino acid residues in metal proteins is site-directed mutagenesis, in which one or more residues are substituted in order to study the influence of this mutation on the functions and properties of the metal protein, like reduction potential[3] and UV-VIS[36,54,81,319,322], EPR[23,38,74,302-304], and NMR[33,36,47,56,76,78,80,84,319-321,327-329] spectra.

In this section, the active site structure of a few metal substituted azurin molecules or azurin mutants for which a crystal structure is available, is optimized using QM/MM calculations. This enables a further check on the reliability of using the QM/MM method for obtaining active site structures, by comparing the calculated and experimental structures. Furthermore, the computed active site structure is used to obtain UV-VIS spectra and EPR data that can be checked with experimental data.

Zn-substituted azurin

In the preparation of azurin samples for use in spectroscopic studies, it was shown that the samples may contain another protein that co-purified with azurin and could not easily be distinguished from the latter.[59] In later studies, using among others new isolation procedures, it was possible to isolate the contaminating by-product, which was called azurin*. It was shown that this protein was highly similar to azurin, but was not the oxidized, reduced or apo-form of it, nor could it readily incorporate copper ions. Using several techniques, it was established that a “heavy” atom was bound in the active site, which was proposed and confirmed to be zinc based on, among other evidence, the absence of any features in the visible absorption spectrum. X-ray crystallographic data[59] showed a distinctly different binding of the metal to the active site residues. Compared to wildtype azurin, the zinc atom moves 0.3 Å down towards Gly45, which moves 0.3 Å upwards itself. Unlike copper, which is found at 0.08 Å above the NNS-plane of His46, His117 and Cys112, zinc is located at 0.16 Å below this plane. As a result, the metal-Gly45 distance is decreased from 2.97 Å for copper to 2.32 Å for zinc.

A hybrid QM/MM calculation was performed to optimize the active site geometry of zinc-azurin, where the starting structure was taken from wildtype copper-azurin, with copper replaced by zinc. The remaining details of the QM/MM setup are similar to the one used for wildtype azurin as described in Section 9.2. The optimized zinc-residue distances are given in Table 9.3.1. Even though in the starting structure the zinc atom is at some 2.85 Å from Gly45, it moves towards it in the QM/MM calculation and at the end is located at a distance of only 2.17 Å from it. The zinc atom is then found 0.16 Å below the NNS plane, just like found by the X-ray data. The distances to the in-plane ligands are very well described, and are found to be slightly larger than for *wildtype* azurin. This feature is observed both in

the X-ray crystallographic studies and the QM/MM calculation for the metal-Cys112 distance, which increases by 0.063 Å (X-ray) or 0.071 Å (QM/MM) in going from copper to zinc.

TABLE 9.3.1. QM/MM OPTIMIZED DISTANCES (Å) OF ZN-AZURIN

	X-ray	*QM/MM*
Zn-Gly45 [a]	2.319 ± 0.091	2.173
Zn-His46 [a]	2.069 ± 0.050	2.085
Zn-Cys112 [a]	2.300 ± 0.022	2.240
Zn-His117 [a]	2.007 ± 0.036	2.056
Zn-Met121 [a]	3.383 ± 0.094	3.484
Zn-N2S [a]	-0.164 ± 0.042	-0.158

a) see subscript Table 9.2.1 for definition

Co/Ni substituted azurin

Recently, also the crystal structures of cobalt[73] and nickel[330] substituted azurin have become available, which showed an active site structure similar to zinc-azurin[59]. The metal atom is found in these three cases in a distorted tetrahedral coordination, where the metal atom is located below the NNS plane. The Met121 residue is found at even larger distances than in *wildtype* azurin. Some remarkable features of the cobalt and nickel azurin structures are however the distances of the metal to the in-plane ligands. Bonander et al.[73] performed a search in the Cambridge Data Base (CDB), looking for metal compounds with the metal coordinated to at least two nitrogens and a sulphur. The average metal-nitrogen distance is for both cobalt and nickel 2.02 Å, with a maximum of 2.11/2.14 Å. However, for cobalt azurin the His residues are found at 2.32 and 2.25 Å, while for nickel azurin they are located at 2.23 and 2.22 Å. The search in the CDB also gave an average nickel-oxygen distance of 2.10 Å, with a maximum of 2.16 Å, and an average cobalt-oxygen distance of 1.98 Å with a maximum of 2.16 Å.

In this case, two separate QM/MM calculations have been performed for both cobalt and nickel azurin. The QM/MM setup was the same as described before for wildtype azurin (Section 9.2). In the first try, the starting structure was taken from *wildtype* azurin and copper was replaced by either cobalt or nickel, while in a second run the starting structure was taken from the cobalt azurin crystal structure. The fully optimized active site structure is in both cases the same, whether started from the *wildtype* or the cobalt-azurin starting structure. The QM/MM optimized metal-residue distances are reported in Table 9.3.2.

Unlike the QM/MM calculations on *wildtype* and zinc azurin, the difference between the calculated and experimental metal-ligand distances is not within the experimental uncertainty anymore. The in-plane ligands are found in all cases closer to the metal atom in the QM/MM calculations than in the X-ray data. The same trend has been observed by de Kerpel et al.,[104] who performed calculations on small model systems of the active site of cobalt-azurin. Their cobalt-His distances are somewhat longer (~0.05 Å) than the ones observed here, but are still significantly smaller than the reported X-ray values[73]. Moreover, the QM/MM cobalt-His distances reported here agree very well with the average Co-N distance (2.02 Å) found in the Cambridge Data Base (CDB), as described above[73]. Since the X-ray Co-N distances are 0.12-0.28 Å longer than the *maximum* observed in the CDB, this may suggest that the cobalt-nitrogen distances observed in the crystal structure of Co-azurin[73] are too long.

TABLE 9.3.2. QM/MM OPTIMIZED DISTANCES (Å) OF CO/NI-AZURIN ACTIVE SITE

	X-ray Co-az.[73]	*QM/MM Co-az.*	*X-ray Ni-az.*[330]	*QM/MM Ni-az.*
Co/Ni-Gly45[a]	2.230 ± 0.077	2.089	2.46 ± 0.06	2.123
Co/Ni-His46[a]	2.243 ± 0.045	2.038	2.23 ± 0.09	2.059
Co/Ni-Cys112[a]	2.337 ± 0.024	2.196	2.39 ± 0.07	2.172
Co/Ni-His117[a]	2.421 ± 0.099	2.013	2.22 ± 0.12	2.055
Co/Ni-Met121[a]	3.561 ± 0.126	3.552	3.30 ± 0.05	3.209
Co/Ni-N2S[a]	-0.207 ± 0.071	-0.230	-0.18	-0.082

a) see Table 9.2.1 for definition

The distance of cobalt from the Cys112 and Gly45 residues is much smaller in the QM/MM calculations than in the X-ray data. The QM/MM calculated distance to Gly45 of 2.09 Å falls again well within the range observed in the CDB, while the X-ray distance of 2.23 Å is 0.07 Å larger than the maximum observed in CDB. Also for Cys112 a much smaller distance is observed in the QM/MM calculations than in the X-ray data. On the other hand, for Met121 a good agreement between calculated and experimental distances is found. In both cases, this residue is located at some 3.55 Å from the metal and can therefore have at most a very weak bonding with the metal atom. Another feature preserved in both QM/MM and X-ray data is the position of the metal with respect to the NNS plane. The cobalt atom is located in both cases below this plane at some 0.21/0.23 Å.

The spin density[a] of cobalt-azurin (amounting a total of three electrons due to the quartet ground state) is delocalized in the active site also. On the metal itself resides a total spin of 1.830 (61.0 %), while a total of 0.498 (16.6 %) is located on the Cys112 sulphur. This is a different situation then for *wildtype* azurin, where the spin density was mainly found on Cys112 and only roughly 20 % on copper. The remaining spin density is located on the Gly45 oxygen (0.133; 4.4 %), the His46 nitrogen (0.092; 3.1 %), the His117 nitrogen (0.121, 4.0 %) and the Met121 sulphur (0.031; 1.0 %). Therefore, even though the cobalt-sulphur distance is rather large (3.55 Å), there is still a weak covalent interaction present. This agrees well with some peaks observed in the 1D ^{1}H NMR spectrum of Co(II)-azurin, which have been assigned to protons in the methyl group of the Met121 residue. These hydrogens carry a small spin density of ~0.006 or 0.2 %. Just like the relative amount of spin density on the sulphur of Cys112 is smaller than in wildtype azurin, also the relative amount of spin density on the β-hydrogens is smaller by a factor of approximately three. Compared to wildtype azurin, a considerable increase of spin density is observed for His46-H$_\alpha$ (1.1 %), while it stays roughly the same for the other hydrogens in His46 and His117 (0.1-0.4 %).

For nickel-azurin, the same pattern is observed as for cobalt-azurin. The nickel-His distances observed in the X-ray studies are ~0.12 Å larger than the *maximum* observed in the CDB, while the QM/MM optimized nickel-His distances of ~2.05 Å agree perfectly with the average distance observed in the database. The same feature is observed for the nickel-Gly45 distance, which is found at 2.46 Å in the X-ray data and at 2.12 Å in the QM/MM calculations. Again, the distance from X-ray data is larger (0.26 Å) than the maximum observed in the database, while the QM/MM distance fits perfectly with the average Ni-O distance in the database. However, for the nickel-Cys112 distance the opposite is found. The distance from X-ray data fits very well with the average distance in the CDB, while the QM/MM optimized distance is found at almost 0.2 Å smaller than the minimum. Still, as there is only

[a] From Multipole Derived Charge analysis; see sections 3.1 and 9.2 for more details

a small difference of less than 0.03 Å with the metal-Cys distance in either wildtype or cobalt-azurin, this preserved metal-Cys112 distance may be typical for azurin. The Met121 is found in both the calculated and experimental structure at a shorter distance from nickel than from cobalt by some 0.2-0.3 Å, while the nickel atom itself is located, just like zinc and cobalt, below the NNS plane of His46, His117 and Cys112, by some 0.1/0.2 Å.

For nickel-azurin, the spin density[a], which amounts to two electrons for the triplet ground state, is found mainly on the nickel (1.01, 50.5 %), Cys112 sulphur (0.46, 23 %), Gly45 oxygen (0.111, 5.5 %), the His46 (0.072, 3.6 %) and His117 (0.094, 4.7 %) nitrogens and the Met121 sulphur (0.060, 3 %). This delocalization of spin on Met121 agrees, just like cobalt-azurin, well with some peaks in NMR spectra, which were assigned to protons in the Met121 residue.

Asn47Leu mutant

The Asn47 residue is preserved in all type 1 copper proteins for which the structure has been resolved so far, which makes it likely that it has a special function in the protein. In fact, it is found that the sidechain of the residue forms hydrogen bonds to Tyr113, which link the loops containing the copper ligands, while the backbone amide nitrogen forms a hydrogen bond towards the sulphur of Cys112. The former hydrogen bonds are presumably important for the stability and rigidity of the protein, while the hydrogen bond to the Cys112 sulphur may be involved in the peculiar spectroscopic features of the type 1 copper proteins.

The Asn47Leu mutant of *Alcaligenes denitrificans* azurin has been made in order to study the importance of this residue for the structure and function of azurin. Characterization of the mutant by UV-visible, electron spin resonance, and NMR spectroscopy revealed that the spectroscopic features are very similar to the ones of wildtype azurin. The only significantly different features are the midpoint reduction potential that changes from 286 mV (*wildtype*) to 396 mV (*N47L mutant*), and the temperature dependence of the midpoint potential, showing distinctly different entropic and enthalpic effects upon reduction. It was shown that the change in reduction potential could be reasonably well reproduced by the electrostatic interaction of the Asn47 sidechain atoms with the *extra* charge that "enters" on copper upon reduction, under the assumption that this *extra* charge resides at the copper atom. However, it has been shown in Section 6.1 that this extra charge delocalizes in the active sites, and resides on copper for only roughly 13 %.

The QM/MM optimized structure of the N47L-azurin mutant has been obtained like described previously, and results in copper-residue distances that are given in Table 9.3.3. The distances to the in-plane ligands are very similar to the *Pseudomonas aeruginosa* wildtype azurin values, but some differences are observed for the axial residues. While the distance of copper from either Met121 or the N2S-plane in the reduced state is similar to the wildtype values, the distance from Gly45 is smaller by some 0.16 Å. This difference is increased in the oxidized state, where not only the copper-Gly45 but also the copper-Met121 distance differs by some 0.3 Å from the wildtype structure. The Gly45 oxygen has moved upwards considerably (~0.2 Å), while the copper moved downwards by 0.12 Å. At the same time, the Met121 sulphur has moved away from the copper by some 0.2 Å. These changes with respect to the *wildtype* structure may have been caused by using the structure from *Alcaligenes denitrificans* azurin instead of *Pseudomonas aeruginosa* azurin; making the N47L change in the *Pa* azurin structure and optimizing it with QM/MM in the oxidized state leads to copper-ligand distances that are virtually identical to the *wildtype Pa* values, with values of 2.803 (Gly45), 2.023 (His46), 2.176 (Cys112), 2.000 (His117) and 3.002 (Met121). The only significant change in comparison with *wildtype* azurin is found

[a] From Multipole Derived Charge analysis; see sections 3.1 and 9.2 for more details.

for the axial groups with changes in the copper-ligand distance of –0.052 (Gly45) and –0.176 (Met121).

TABLE 9.3.3. QM/MM OPTIMIZED DISTANCES (Å) OF ASN47LEU AZURIN ACTIVE SITE

	X-ray[280] *reduced state*	*QM/MM reduced state*	*X-ray*[280] *oxidized state*	*QM/MM oxidized state*
Cu-Gly45 [a]	3.264 ± 0.107	2.840	3.075 ± 0.175	2.507
Cu-His46 [a]	2.047 ± 0.097	2.035	1.947 ± 0.039	1.999
Cu-Cys112 [a]	2.153 ± 0.102	2.227	2.139 ± 0.013	2.168
Cu-His117 [a]	2.211 ± 0.084	2.069	2.047 ± 0.147	1.989
Cu-Met121 [a]	3.091 ± 0.067	3.342	3.087 ± 0.047	3.479
Cu-N2S [a]	0.088 ± 0.041	0.071	0.052 ± 0.002	0.036

a) see subscript Table 9.2.1 for definition

Still, in the QM/MM optimized *Alcaligenes denitrificans* structure, the coordination of the strong in-plane ligands remains preserved, including the hydrogen bond of the amide nitrogen of Leu47 to the Cys112 sulphur (at 3.39 Å), as well as the hydrogen bond of the Phe114 amide nitrogen to the sulphur (3.91 Å). Therefore, although the axial structure gets distorted considerably, the more important in-plane ligands remain close to their wildtype positions. This would account for the almost negligible effect of the mutation on the spectroscopic features, while the difference in the axial residue positions might account for the difference in reduction potential and the entropic and enthalpic effects on these.

Asn47Asp mutant

Whereas the Asn47Leu mutant described above did not change the charge of residue 47, the Asn47Asp (N47D) mutation[4] of *Pseudomonas aeruginosa* (*Pa*) azurin introduces a negative charge close to the copper site. The influence of this introduced charge on the stability and functions of the protein has been investigated by X-ray crystallographic, spectroscopic and kinetic studies[4]. The crystal structure was obtained for the zinc mutant[4], and was found to be very similar to the one of wildtype zinc-azurin[59]. Unlike the N47L-mutant that had been purified to get rid of the zinc-substituted form, the N47D-mutant was obtained with a copper content of only 3 %. However, zinc is spectroscopically silent in the visible region of the absorption spectra as well as in EPR spectra, and because of its electronic properties does not show any reduction potential. Therefore, the spectroscopic properties of the copper-mutant could be determined. The spectroscopic and kinetic properties are similar to wildtype *Pa* azurin, while a reduction potential is found that is only 20 mV higher.

This mutant poses a challenge for the QM/MM method, as the N47D-structure for copper has not yet been determined. Therefore, the active site structure was optimized not only for the zinc N47D mutant but also for the copper-protein. The QM/MM optimized metal-residue distances are given in Table 9.3.4. Just like observed for the N47L mutant, the coordination of the in-plane ligands around the metal atom is very similar to *wildtype* azurin, thereby accounting for the negligible effect of the mutation on the spectroscopic features. The coordination of the axial residues is for the zinc-N47D structure rather similar to the wildtype zinc structure, although the metal has shifted downwards by ~0.1 Å.

TABLE 9.3.4. QM/MM OPTIMIZED DISTANCES (Å) OF ASN47ASP ZN/CU AZURIN ACTIVE SITE

	X-ray[4] *Zn-N47D*	*QM/MM Zn-N47D*	*X-ray*[7] *Cu-wildtype*	*QM/MM Cu-N47D*
Zn/Cu-Gly45 [a]	2.358 ± 0.055	2.167	2.967 ± 0.093	2.536
Zn/Cu-His46 [a]	2.096 ± 0.054	2.100	2.076 ± 0.060	2.003
Zn/Cu-Cys112 [a]	2.271 ± 0.068	2.266	2.237 ± 0.044	2.160
Zn/Cu-His117 [a]	2.047 ± 0.144	2.056	2.011 ± 0.069	2.004
Zn/Cu-Met121 [a]	3.439 ± 0.060	3.557	3.149 ± 0.070	3.307
Zn/Cu-N2S [a]	-0.136 ± 0.035	-0.264	0.083 ± 0.051	0.028

a) see subscript Table 9.2.1 for definition

The axial coordination in the copper N47D structure is, like was observed for the N47L structure, different from the wildtype structure. The distance of the Gly45 oxygen to copper is decreased by roughly 0.3 Å to 2.54 Å, while the distance of the Met121 sulphur from copper is increased by some 0.13 Å to 3.31 Å. This is accompanied by a downward move of copper itself of some 0.06 Å.

The hydrogen bonds towards the Cys112 sulphur are either longer (Asp47 backbone nitrogen) or shorter (Phe114 backbone nitrogen) by roughly 0.2 Å in this mutant. The mutation doubles the number of hydrogen bonds between the 47 and 113 residues. The backbone nitrogen of Thr113, which is hydrogen bonded to the sidechain oxygen of Asn47 in the wildtype structure, is now connected to both sidechain oxygens of Asp47. This is accomplished by a rotation of the Asp47 sidechain to reduce the distance of the second oxygen from the nitrogen. A fourth hydrogen bond is observed between the second oxygen and the backbone nitrogen of Phe114.

Phe114Ala mutant

The Phe114 residue is part of a hydrophobic patch at the northern end of azurin near the copper site and exhibits a π-electron overlap with the copper ligand His117, which has been suggested to be involved in the electron self exchange (e.s.e.) rate of the protein[6,7,95]. To study this suggested role the Phe114Ala mutant of *Pa* azurin had been made[5], which creates a hole in the immediate surroundings of the copper site. This mutation was found to have a non-negligible effect on the spectroscopic features of the protein; the wavelength of maximum absorption in the UV-vis spectra decreases by ~10 nm with an intensity that is only 78 % compared to wildtype *Pa* azurin, the hyperfine splitting parameter increases by 6 % and the reduction potential increases by ~50 mV. The latter may be not that important, as also the N47D and N47L mutants showed reduction potentials that differ from the wildtype one. The crystal structure (at 2.6 Å resolution) showed an active site structure that is similar to wildtype azurin, except for the copper-His distances which are elongated by a small amount in the case His46 (2.13 Å) and by a large amount for His117 (2.46 Å). It is uncertain whether this is due to the mutation or to the uncertainty imposed by the resolution of the crystal.

Therefore, the geometry of the active site was optimized in the presence of the protein using hybrid QM/MM calculations. The QM/MM optimized copper-residue distances are given in Table 9.3.5, which differ quite a bit from the wildtype values.

TABLE 9.3.5. QM/MM OPTIMIZED DISTANCES (Å) OF PHE114ALA AZURIN ACTIVE SITE

	X-ray reduced state	*QM/MM reduced state*	*X-ray*[5] *oxidized state*	*QM/MM oxidized state*
Cu-Gly45 [a]	-	2.867	2.983 ± 0.138	2.555
Cu-His46 [a]	-	2.075	2.127 ± 0.071	2.037
Cu-Cys112 [a]	-	2.319	2.228 ± 0.012	2.185
Cu-His117 [a]	-	2.134	2.463 ± 0.128	2.003
Cu-Met121 [a]	-	3.061	3.022 ± 0.121	3.506
Cu-N2S [a]	-	0.173	0.431 ± 0.289	-0.022

a) see subscript Table 9.2.1 for definition

The in-plane coordination is rather similar, but just like found for the N47L and N47D mutants, the Gly45 oxygen moves closer to copper by some 0.3 Å. At the same time, the distance of the Met121 sulphur from the copper is increased by almost the same amount. A rather peculiar conformation for the Met121 residue is observed in this mutant. The distance of the Cβ of Met121 from the Cβ of Cys112 (normally found between 3.65 and 3.8 Å) decreases to 3.41 Å and is accompanied by a $C\gamma_{121}$-$C\beta_{112}$ distance of only 3.98 Å (normally 4.45-4.55 Å). This is not observed in either the N47D or the N47L mutants, which showed normal values for these distances.

The hydrogen bond from the backbone nitrogen of residue 114 to the Cys112 sulphur is found at a shorter distance of 3.48 Å, while the Asn47 backbone nitrogen to Cys112 sulphur hydrogen bond is elongated at 3.67 Å. This feature was also present in the N47D mutant and can therefore not be responsible for the change in spectroscopic features.

Met121Gln mutant

The Met121Gln mutant has received some special attention[36-38,42,86,90,103], as it has been used as model for stellacyanin, a type 1 copper protein for which the structure has been elucidated only recently[96]. It had been proposed that the fourth ligand in stellacyanin might be glutamine, and the spectroscopic features of the M121Q azurin mutant[42] were indeed very similar to the ones observed in stellacyanin. The crystal structure of the oxidized M121Q mutant shows a tetrahedral coordination around copper, with the usual three in-plane ligands (His46, Cys112, and His117) and Gln121 as the fourth ligand at 2.26 Å. The fifth coordinating residue that is normally found in azurin (Gly45) is located at 3.37 Å in the crystal structure at a weakly bonding position. Upon reduction, a peculiar crystal structure is observed; the distances of copper to Gly45, His46 and Cys112 remain rather the same, but the Gln121 and His117 residues move away from copper to 2.73 and 2.67 Å respectively. Also the angles around the copper change dramatically, thereby changing the coordination from approximately tetrahedral to almost linear with two strong ligands (His46 and Cys112).

The active site structure of both the reduced and oxidized form of the M121Q-mutant of *Alcaligenes denitrificans* (*Ad*) azurin has been optimized in the presence of the protein and a layer of solvent molecules surrounding the protein, as described for wildtype *Pa* azurin (Section 9.2). The QM/MM optimized copper-residue distances are given in Table 9.3.6.

TABLE 9.3.6. QM/MM OPTIMIZED DISTANCES (Å) OF MET121GLN AZURIN ACTIVE SITE

	X-ray[42] *reduced state*	*QM/MM reduced [p/dp]*	*X-ray*[42] *oxidized state*	*QM/MM oxidized state*
Cu-Gly45 [a]	3.332 ± 0.089	3.002 / 2.862	3.372 ± 0.136	3.341
Cu-His46 [a]	1.975 ± 0.062	1.885 / 2.004	1.936 ± 0.041	1.950
Cu-Cys112 [a]	2.089 ± 0.030	2.114 / 2.221	2.121 ± 0.019	2.148
Cu-His117 [a]	2.675 ± 0.078	3.086 / 2.074	2.046 ± 0.025	2.022
Cu-Gln121 [a]	2.726 ± 0.051	3.086 / 3.327	2.263 ± 0.017	2.300
Cu-N2S [a]	0.200 ± 0.004	0.039 / -0.057	0.260 ± 0.007	0.270

p) His117 protonated at δ position

dp) His117 deprotonated at δ position

a) see Table 9.2.1 for definition

At first, the geometry of the reduced protein was optimized using the normal protonation states of the residues in the site, e.g. both His46 and His117 were protonated at the ε position and bonded to copper at the δ position (*dp* entries in Table 9.3.6). However, this results in an active site structure similar to the reduced wildtype structure, with a distance of the copper to His117 that is more than 0.6 Å smaller than in the crystal structure. This deviation is too large to be accounted for by the experimental uncertainty, especially since the resolution of the crystal structure is 1.94 Å. Therefore, some other effect must be responsible for this peculiar structure. As a second attempt it was checked whether His117 could have been protonated at both nitrogen positions. The results are also given in Table 9.3.6 (*p* entries), which give a somewhat better agreement with the crystal structure regarding the distances of copper to Cys112 and His46, but not for the distances to His117 and Gln121 that are 0.41 and 0.36 Å too long respectively. Given the resolution of 1.94 Å, a superposition of the *p* and *dp* results is also ruled out. Finally, it has been checked whether the His117 residue might have been rotated as to orient with the ε nitrogen towards the copper. Both the deprotonated and protonated nitrogen structure (now at the ε position) has been optimized, which resulted for the protonated case in an almost identical structure as the one where both nitrogens were protonated and the copper was connected to the δ nitrogen. The structure of the deprotonated ε-nitrogen was found to give even longer copper-His117 distances. As these four QM/MM calculations all resulted in structures that disagree to a large extent with the crystal structure, it seems that the crystal structure does not correspond to either one of the systems optimized by the QM/MM calculations.

The oxidized protein structure is very well reproduced by the QM/MM calculations. The difference between the crystal and QM/MM optimized structure is in all cases very small.

Co/Ni substituted Met121Gln mutant

Just like for wildtype azurin, also for the M121Q mutant has copper been replaced by cobalt and nickel to study the spectroscopic features of the mutant with paramagnetic proton NMR[36]. However, it is shown earlier in this section that the metals may coordinate in different manners. The extent with which they do in the M121Q mutant has also been studied using hybrid QM/MM geometry optimizations of the active site in presence of the protein. The optimized copper-residue distances are given in Table 9.3.7, where for reference also the QM/MM optimized and crystal structure data of the copper M121Q protein are given.

TABLE 9.3.7. QM/MM OPTIMIZED DISTANCES (Å) OF CO/NI MET121GLN AZURIN ACTIVE SITE

	X-ray[42] *Cu*	*QM/MM Cu*	*QM/MM Co*	*QM/MM Ni*
Co/Ni-Gly45 [a]	3.372 ± 0.136	3.341	2.792	3.079
Co/Ni-His46 [a]	1.936 ± 0.041	1.950	2.015	2.002
Co/Ni-Cys112 [a]	2.121 ± 0.019	2.148	2.189	2.158
Co/Ni-His117 [a]	2.046 ± 0.025	2.022	1.998	2.009
Co/Ni-Gln121 [a]	2.263 ± 0.017	2.300	2.158	2.068
Co/Ni-N2S [a]	0.260 ± 0.007	0.270	0.247	0.258

a) see subscript Table 9.2.1 for definition

For both cobalt and nickel M121Q azurin, the position of the metal relative to the N2S plane is reversed relative to wildtype cobalt/nickel substituted (*Co-Ni wt*) azurin; this is due to the axial bonding of the metals that takes place in M121Q azurin for both metals with Gln121 instead of Gly45, as observed in *Co-Ni wt* azurin. The distances to the in-plane ligands remain almost the same as in *Co-Ni wt* azurin, while the bond to oxygen is shorter than in *Co-Ni wt* azurin for nickel and longer for cobalt by some 0.05-0.07 Å. The distance to the other axial ligand, in this case Gly45, is shorter than in *Co-Ni wt* azurin for both metals by 0.76 Å (Co) and 0.13 Å (Ni) respectively.

Compared with the copper-M121Q structure a larger similarity is observed between the three metals than for *Co-Ni wt* azurin. All three metal sites show more or less a tetrahedral coordination, with only a weakly bonding interaction with Gly45.

Conclusions

The effect of metal substitution and and/or residue mutations can be very well reproduced by hybrid QM/MM geometry optimizations, in which the active site structure is optimized in the presence of the protein and solvent molecules. Even when starting the QM/MM calculations with structures that are not biased for the metal in the site, the metal atom moves away towards its favored position.

The zinc/cobalt/nickel substituted azurin structures show a tetrahedral coordination around the metal, as shown also by the crystal structures. In these metal sites, the metal atom moves below the N2S-plane of the strong in-plane ligands (His46, Cys112 and His117), thereby getting closer to the Gly45 oxygen. The latter residue moves upwards itself, leading to a metal-oxygen distance in a range of 2.09-2.17 Å for the three metals. The distance of the His46, His117 and Gly45 residues from either cobalt or nickel is shorter than observed in the crystal structure, at distances that are normal for these kind of bonds.

The function of the Asn47 residue, which is conserved in all type 1 copper proteins for which the structure is known, had been studied by site directed mutagenesis by mutating it to either Leu (N47L) or Asp (N47D). For both these mutants, the active site structure has been optimized with QM/MM calculations; for the N47D mutant, not only the copper-protein has been used but also the zinc-structure, as the crystal structure of it was obtained for the latter. In the N47L mutant, two hydrogen bonds connecting residue 47 with Thr113 are removed, which may have impact on the stability of the protein. However, the hydrogen bond of the backbone nitrogen towards the Cys112 sulphur is preserved in both the N47L and the N47D mutant. In the latter, instead of removing two hydrogen

bonds, two are added between the residues 47 and 113. This is brought about by a rotation of the Asp47 sidechain to facilitate these two extra hydrogen bonds. In both cases, the Gly45 oxygen moves closer towards the copper at a distance of ~2.5 Å. This may have an effect on the reduction potential, but as the coordination of the strong in-plane ligands and the hydrogen bonds towards the Cys112 sulphur are preserved, the spectroscopic features hardly change.

The Met121Gln mutant can in the oxidized state be very well described by the QM/MM calculations; the differences between the QM/MM optimized and crystal structure are small. In the reduced state however, the crystal structure can not be reproduced by the QM/MM calculations. In this peculiar crystal structure, the His117 residue is found at some 2.7 Å from the copper. Neither the normal situation, with His117 connected to copper with its *deprotonated* Nδ atom, nor an unusual one, with His117 connected to copper with the *protonated* Nδ atom, provides an active site similar to the crystal structure. In the first, the His117 residue is found at its normal position at some 2.05 Å from the copper, while in the second it is found at a too large distance of 3.1 Å. It has also been checked whether the residue might have rotated 180°, thereby connecting to the copper with its Nε atom, but this resulted in a structure similar to the previous *protonated* His117 one, therefore different from the crystal structure. Therefore, it seems that the crystal structure does not correspond to either one of the systems optimized by the QM/MM calculations.

9.4 *H117G/N42C azurin*

An engineered double mutant exhibiting unexpected spectroscopic features[a]

The His117Gly mutant[75,319,324,325] creates a hole in the protein surface through which exogenous ligands can enter the active site and bind to the copper. Depending on the kind of ligand, different types of copper sites are obtained; binding of chloride, bromide, imidazole (and derivatives) or pyridine (and derivatives) restores the characteristic spectroscopic features of a type 1 copper site[319], whereas the binding of for instance water or histidine creates novel azurin copper sites, which are more similar to "normal" (or type 2) copper sites. For instance, without external ligands, His117Gly azurin is green with absorption maxima at 420 and 628 nm, but addition of imidazole restores the characteristic blue color of wildtype azurin.

In an attempt to construct dimers of *Pseudomonas aeruginosa* azurin, the double mutant H117G/N42C has been made[61] that would enable a dimer by the formation of a sulphur bond between both Cys42's, which has been successfully applied in the N42C single mutant[327] and was shown to be catalyzed by the presence of copper. Adding Cu(II) to the colorless monomeric form of apo-H117G/N42C results in the immediate formation of a yellow color, with a strong absorption at 385 nm[61]. The addition of external ligands like imidazole or chloride does not influence the spectroscopic features, therefore the open binding site on copper due to the His117Gly must be occupied by another ligand. The most likely candidate for this new ligand is then Cys42, which was proven by several experiments; for instance, adding Cu(II) to the apo-H117G/N42C dimer results in the expected green color of the H117G mutant, while the blocking of Cys42 by iodoacetamide also restores the characteristic properties of the H117G mutant.

The absorption at 385 nm is characteristic of a type-2 copper site and can be assigned as a cysteine sulphur to copper charge transfer band. In the EPR spectrum of the double mutant two species can be distinguished, both with tetragonal type-2 copper site characteristics, which can be attributed to intermediates that are formed during the uptake of copper. After three days, only the most stable species remains with typical type-2 site characteristics: a g_z value of 2.23 with a hyperfine splitting of 0.0160 cm^{-1}.

Molecular Dynamics simulations

As it has been suggested that the Cys42 residue is picking up the copper from the solution and bringing it into the active site, the same process has been investigated with Molecular Dynamics simulations. At the beginning the copper was bonded to Cys42 and then brought into the active site by applying distance restraints in six steps. Inbetween any of these steps, the protein was allowed to react to the new situation and relax. As the experimental results indicated a tetragonal coordination around copper, in the last step an improper dihedral potential has been added to ensure a tetragonal coordination. At the end of the process, the copper was present in the active site at 2.35 Å from His46, 2.45 Å from Cys112, 2.48 Å from Cys42 and 2.81 Å from Met121. These copper residue distances show a reasonable amount of strain on the residues, which might have been build up in the process of bringing the copper into the active site.

[a] Joint study by M. van den Bosch (MD simulations) and M. Swart (QMMM optimizations)

QM/MM geometry optimization

Starting from the structure obtained in the MD simulation, the active site geometry has been optimized in the presence of the protein and a surrounding layer of water molecules using hybrid QM/MM calculations, which were setup as described in Section 9.2. The optimized copper-residue distances are given in Table 9.4.1.

TABLE 9.4.1. DISTANCES (Å) IN H117G/N42C-AZURIN ACTIVE SITE

	Molecular Dynamics	*QM/MM*	*QM/MM (+H2O)*
Cu-Cys42	2.48	2.29	2.31
Cu-Gly45	5.38	5.32	4.99
Cu-His46	2.35	2.04	2.05
Cu-Cys112	2.45	2.18	2.21
Cu-Met121	2.81	4.01	3.91
Cu-H2O	-	-	2.89

The strain in the copper site has been removed, as indicated by the copper-His46 and copper-Cys112 distances that are at normal values again, while also Cys42 has moved closer to the copper at 2.29 Å. The Met121 on the other hand has completely moved away from the copper site and is not involved as the proposed fourth tetragonal ligand, nor does any other ligand move closer to take its place.

A possible fourth ligand for the copper in this site might be a water molecule that could have stayed connected to copper during the picking up and subsequently bringing in into the active site of copper, or it may already been present in the active site. Therefore, a water molecule was placed in the active site close to the space occupied by Met121 before it moved away, at roughly 2.0 Å from the copper, which is a normal distance for a copper-oxygen bond. In the QM/MM geometry optimization however, the water moves away from the copper just as Met121 did, and is located at 2.9 Å in the end. The ligands remain approximately at the same position, while Gly45 and Met121 move a little closer to the copper again. Still, the coordination is far from tetragonal and therefore also the water molecule can not be the fourth ligand.

A confirmation that the strain in the protein has been removed in the QM/MM optimization is given by the fact that the protein remained stable in a Molecular Dynamics simulation, that was performed on the QM/MM optimized structure without a water molecule in the active site, and resulted in normal copper residue distances again. As this strain was not released in the intermediate relaxation parts of the initial MD simulations, it indicates that the bringing in of the copper may have been done too fast and with too hard pulling. Currently, this process is reinvestigated with MD simulations where the pulling in of the copper is performed at a lower speed and with less brute force.

Computation of EPR parameters

The g-tensor and hyperfine splitting for the active site obtained in the QM/MM optimization were computed using the experience obtained in *Chapter* 7 (copper nuclear charge of 28.6, in a DZV basis set). The g-tensor values of the double mutant without a water molecule in the active site are 2.025, 2.030, 2.132 with a hyperfine splitting (A_Z) of –398 MHz, while they are 2.027, 2.030, 2.121 with a hyperfine splitting (A_Z) of –360 with a water molecule in the site. These are all considerably lower than the experimental values of 2.23 (g_Z) and -480 MHz (A_Z).[61]

Conclusions

The engineered N42C/H117G double mutant of azurin, which has some remarkable spectroscopic features, has been modeled using a combination of Molecular Dynamics simulations and QM/MM geometry optimizations.

In the simulations, the copper has been attached to Cys42 and pulled inside the active site using distance restraints. Subsequently, the active site structure has been optimized in the presence of the protein, where the presumed four copper ligands (Cys42, His46, Cys112 and Met121) have been put in the QM system. However, in the QM/MM optimization the Met121 immediately drifts away from the copper and ends up at nearly 4 Å, therefore at a non-bonding position.

As the EPR spectra obtained experimentally for the protein seem to indicate that the protein contains a type 2 copper site with four ligands, a water molecule has explicitly been included in the QM system, approximately at the position left open by the Met121 residue, and a new QM/MM geometry optimization performed. However, also the water molecule moves away from the copper and can not be the purported fourth ligand.

The computed g-tensor values of the optimized active site structures are too small in comparison with the experimental values, thereby indicating that the optimized structures, with copper effectively bonded to three ligands, can not be responsible for the spectroscopic features.

APPENDICES

A

CONTENTS

SUMMARY

A description is given of the units used in this thesis, and a list of frequently used abbreviations with the corresponding term is given.

A.1 *Units*

Description of units used in this thesis and conversion factors for transformation into other units

The formulas and properties presented in this thesis are reported in *atomic units* unless explicitly noted otherwise; the exceptions to this rule are energies, which are most frequently reported in kcal/mol, and distances that are normally reported in Å.

In the *atomic units* system, four frequently used quantities (Planck's constant h divided by 2π [$\hbar$], mass of electron [m_e], electron charge [e], and vacuum permittivity [$4\pi\varepsilon_0$]) are set explicitly to 1 in the formulas, making these more simple to read. For instance, the Schrödinger equation for the hydrogen atom is in SI units:

$$\left(-\frac{\hbar^2}{2m_e}\nabla^2 - \frac{e^2}{4\pi\varepsilon_0 r}\right)\psi = E\psi \tag{1}$$

In atomic units, it looks like:

$$\left(-\frac{1}{2}\nabla^2 - \frac{1}{r}\right)\psi = E\psi \tag{2}$$

Before a quantity can be used in the *atomic units* equations, it has to be transformed from SI units into atomic units; the same is true for the quantities obtained from the equations, which can be transformed from atomic units into SI units. For instance, the solution of equation (2) for the ground state of the hydrogen atom gives an energy of –0.5 atomic units (Hartree), which can be converted into other units quite simply by multiplying with the appropriate conversion factor (see table A.1.1).

Not only are the equations easier to comprehend, it is also easier to apply them on atoms and molecules; suppose that, for instance, the electrostatic interaction between two atoms with unit point charge at 100 Bohr distance from each other is needed. In SI units, one would have to multiply $1/4\pi\varepsilon_0$ twice with the charge in SI units ($1.602 \cdot 10^{-19}$ C), and then divide by the distance (in m!) to obtain the energy in J, which would have to be transformed into kcal/mol. In atomic units it is simply dividing the unit charges by the distance, leading to 0.01 Hartree, or 6.2751 kcal/mol.

Two other useful factors concerning the conversion from energy (in eV) to wavelength (in nm) and are the following:

– wavelength λ (nm) from energy E (eV) $\qquad \lambda = {}^{1239.84}/_{E}$ (3)

– electron resonant frequency ν $\qquad \nu(cm^{-1}) = \nu(MHz) \cdot 0.3335 \cdot 10^{-4}$ (4)

TABLE A.1.1. ATOMIC UNIT SYSTEM AND CONVERSION FACTORS

	quantity	*value in a.u.*	*value in SI units*	
m_e	electron mass	1	$9.10938188(72)\cdot10^{-31}$	kg
$\hbar$	angular momentum	1	$1.054571596(82)\cdot10^{-34}$	Js
h	Planck's constant	2π	$6.62606876(52)\cdot10^{-34}$	Js
e	charge	1	$1.602176462(63)\cdot10^{-19}$	C
$4\pi\varepsilon_0$	vacuum permittivity	1	$1.112650056\cdot10^{-10}$	F/m
a_0	length (Bohr)	1	$5.291772083(19)\cdot10^{-11}$	m
E_h	energy (Hartree)	1	$4.35974381(34)\cdot10^{-18}$	J
			627.50947(5)	kcal/mol
			2625.4996(2)	kJ/mol
			27.2113834(11)	eV
c	speed of light	137.03599976(50)	$2.99792458\cdot10^{8}$ [a]	m
μ	unit dipole moment	1	$8.47835267365(33)\cdot10^{-30}$	Cm
			2.54174619(10)	Debye
α' [b]	unit polarizability	1	$1.481847093(16)\cdot10^{-31}$	m^3
			0.1481847093(16)	$Å^3$
μ_B	Bohr magneton	1/2	$9.27400899(37)\cdot10^{-24}$	J/T
μ_N	nuclear magneton	$2.72308512(11)\cdot10^{-4}$	$5.05078317(20)\cdot10^{-27}$	J/T

a) Exact!
b) In SI units this corresponds to a polarizability volume ($\alpha / 4\pi\varepsilon_0$)

A.2 *Abbreviations*

A list of descriptions for abbreviations that are frequently used in this thesis

Ad	*Alcaligenes denitrificans (azurin)*
ADF	Amsterdam Density Functional (program)
AMBER	Assisted Model Building with Energy Refinement
az.	*azurin*
BSSE	Basis Set Superposition Error
CIS	Configuration Interaction using Singles excitations only
CNDO	Complete Neglect of Differential Overlap
DFT	Density Functional Theory
DRF	Direct Reaction Field (approach)
DZV	Double Zeta Valence (basis set)
EPR	Electron Paramagnetic Resonance
ESR	Electron Spin Resonance
F114A	Phe114 → Ala114 mutant (*azurin*)
GROMACS	GROningen MAchine for Chemical Simulations
GROMOS	GROningen MOlecular Simulation (package)
H117G	His117 → Gly117 mutant (*azurin*)
HF	Hartree-Fock
IMOMM	Integrated Molecular Orbital and Molecular Mechanics (model)
INDO	Intermediate Neglect of Differential Overlap
M121H	Met121 → His121 mutant (*azurin*)
M121Q	Met121 → Gln121 mutant (*azurin*)
MD	Molecular Dynamics (simulations)
MDC-q	Multipole Derived Charges, reproduced up to quadrupole level
MM	Molecular Mechanics
N42C/H117G	Asn42 → Cys42, His117 → Gly117 double mutant (*azurin*)
N47D	Asn47 → Asp47 mutant (*azurin*)
N47L	Asn47 → Leu47 mutant (*azurin*)
ONIOM	Our own N-layered Integrated molecular Orbital + molecular Mechanics (method)
Pa	*Pseudomonas aeruginosa (azurin)*
QC	Quantum Chemistry
QM	Quantum Mechanics
RHF	Restriced Hartree Fock
ROHF	Restriced Open shell Hartree Fock
TD-DFT	Time-Dependent Density Functional Theory
TZ2P	Triple Zeta valence (basis) including double Polarization functions
TZP	Triple Zeta valence (basis set) including Polarization functions
xc	exchange-correlation (potential)
wt	*wildtype* (azurin)

References

1. F. Jensen. *Introduction to Computational Chemistry*, Wiley & Sons, 1998.
2. W.E.B. Shepard, B.F. Anderson, D.A. Lewandoski, G.E. Norris and E.N. Baker. *J.Am.Chem.Soc.* **112** (1990) 7817-7819.
3. T. Pascher, B.G. Karlsson, M. Nordling, B.G. Malmström and T. Vänngård. *Eur.J.Biochem.* **212** (1993) 289-296.
4. L. Sjölin, L. Tsai, V. Langer, T. Pascher, G. Karlsson, M. Nordling and H. Nar. *Acta Cryst.* **D49** (1993) 449-457.
5. L. Tsai, L. Sjölin, V. Langer, T. Pascher and H. Nar. *Acta Cryst.* **D51** (1995) 168-176.
6. E.N. Baker. *J.Mol.Biol.* **203** (1988) 1071-1095.
7. H. Nar, A. Messerschmidt, R. Huber, M. van de Kamp and G.W. Canters. *J.Mol.Biol.* **221** (1991) 765-772.
8. D.S. Bendall. *Protein Electron Transfer*, BIOS Scientific, 1996.
9. K. Pierloot, J.O.A. de Kerpel, U. Ryde and B.O. Roos. *J.Am.Chem.Soc.* **119** (1997) 218-226.
10. K. Pierloot, J.O.A. de Kerpel, U. Ryde, M.H.M. Olsson and B.O. Roos. *J.Am.Chem.Soc.* **120** (1998) 13156-13166.
11. M.H.M. Olsson, U. Ryde and B.O. Roos. *Prot.Science* **7** (1998) 2659-2668.
12. U. Ryde, M.H.M. Olsson, K. Pierloot and B.O. Roos. *J.Mol.Biol.* **261** (1996) 586-596.
13. J.O.A. de Kerpel and U. Ryde. *Prot.Struct.Func.Gen.* **36** (1999) 157-174.
14. M.H.M. Olsson, U. Ryde, B.O. Roos and K. Pierloot. *J.Biol.Inorg.Chem.* **3** (1998) 109-125.
15. G. Battistuzzi, M. Borsari, L. Loschi, F. Righi and M. Sola. *J.Am.Chem.Soc.* **121** (1999) 501-506.
16. R. Lontie. *Copper Proteins and Copper Enzymes*, CRC Press, 1984.
17. U. Ryde and M.H.M. Olsson. *Int.J.Quant.Chem.* **81** (2001) 335-347.
18. B.G. Malmström and P. Wittung-Stafshede. *Coord.Chem.Rev.* **185-186** (1999) 127-140.
19. A.E. Mark and W.F. van Gunsteren. *J.Mol.Biol.* **240** (1994) 167-176.
20. U. Ryde, M.H.M. Olsson, B.O. Roos, J.O.A. de Kerpel and K. Pierloot. *J.Biol.Inorg.Chem.* **5** (2000) 565-574.
21. C. Buning and P. Comba. *Eur.J.Inorg.Chem.* (2000) 1267-1273.
22. P. Comba. *Coord.Chem.Rev.* **200-202** (2000) 217-245.
23. M. van Gastel, G.W. Canters, H. Krupka, A. Messerschmidt, E.C. de Waal, G.C.M. Warmerdam and E.J.J. Groenen. *J.Am.Chem.Soc.* **122** (2000) 2322-2328.
24. A. Luise, M. Falconi and A. Desideri. *Prot.Struct.Func.Gen.* **39** (2000) 56-67.
25. S. DeBeer, P. Wittung-Stafshede, J. Leckner, G. Karlsson, J.R. Winkler, H.B. Gray, B.G. Malmström, E.I. Solomon, B. Hedman and K.O. Hodgson. *Inorg.Chim.Acta* **297** (2000) 278-282.
26. A. Paciaroni, A.R. Bizzarri and S. Cannistraro. *Phys. B* **269** (1999) 409-415.
27. R. Guzzi, L. Sportelli, C.l. Rosa, D. Milardi, D. Grasso, M.Ph. Verbeet and G.W. Canters. *Biophys.J.* **77** (1999) 1052-1063.
28. A.P. Kalverda, M. Ubbink, G. Gilardi, S.S. Wijmenga, A. Crawford, L.J.C. Jeuken and G.W. Canters. *Biochem.* **38** (1999) 12690-12697.
29. A. Paciaroni, M.E. Stroppolo, C. Arcangeli, A.R. Bizzari, A. Desideri and S. Cannistraro. *Eur.Biophys.J.* **28** (1999) 447-456.
30. X. Wang, S.M. Berry, Y. Xia and Y. Lu. *J.Am.Chem.Soc.* **121** (1999) 7449-7450.
31. X. Wang, M.C. Ang and Y. Lu. *J.Am.Chem.Soc.* **121** (1999) 2947-2948.
32. A.A. Stuchebrukhov. *Int.J.Quant.Chem.* **77** (2000) 16-26.
33. I. Bertini, C.O. Fernandez, B.G. Karlsson, J. Leckner, C. Luchinat, B.G. Malmstrom, A.M. Nersissian, R. Pierattelli, E. Shipp, J.S. Valentine and A.J. Vila. *J.Am.Chem.Soc.* **122** (2000) 3701-3707.
34. E.I. Solomon, D.W. Randall and T. Glaser. *Coord.Chem.Rev.* **200-202** (2000) 595-632.
35. E. Danielsen, R. Bauer, L. Hemmingsen, M.-L. Andersen, M.J. Bjerrum, T. Butz, W. Tröger, G.W. Canters, C.W.G. Hoitink, G. Karlsson, O. Hansson and A. Messerschmidt. *J.Biol.Chem.* **270** (1995) 573-580.
36. J. Salgado, H.R. Jiménez, J.-M. Moratal, S. Kroes, G.C.M. Warmerdam and G.W. Canters. *Biochem.* **35** (1996) 1810-1819.
37. R.E.M. Diederix, G.W. Canters and C. Dennison. *Biochem.* **39** (2000) 9551-9560.
38. J.W.A. Coremans, O.G. Poluektov, E.J.J. Groenen, G.C.M. Warmerdam, G.W. Canters, H. Nar and A. Messerschmidt. *J.Phys.Chem.* **100** (1996) 19706-19713.
39. A.C. Rosenzweig. *Acc.Chem.Res.* **34** (2001) 119-128.
40. A. Chowdhury, L.A. Peteanu, M.A. Webb and G.R. Loppnow. *J.Phys.Chem. B* **105** (2001) 527-534.

41. U. Ryde, M.H.M. Olsson, B.O. Roos and A.C. Borin. *Theor.Chem.Acc.* **105** (2001) 452-462.
42. A. Romero, C.W.G. Hoitink, H. Nar, R. Huber, A. Messerschmidt and G.W. Canters. *J.Mol.Biol.* **229** (1993) 1007-1021.
43. B. Börger, J. Gutschank, D. Suter, A.J. Thomson and S.J. Bingham. *J.Am.Chem.Soc.* **123** (2001) 2334-2339.
44. C. Arcangeli, A.R. Bizzarri and S. Cannistraro. *Biophys.Chem.* **78** (1999) 247-257.
45. C. Arcangeli, A.R. Bizzarri and S. Cannistraro. *Biophys.Chem.* **90** (2001) 45-56.
46. I. Bertini, S. Ciurli, A. Dikiy, C.O. Fernàndez, C. Luchinat, N. Safarov, S. Shumilin and A.J. Vila. *J.Am.Chem.Soc.* **123** (2001) 2405-2413.
47. L. Ma, E. Philipp and J.J. Led. *J.Biomol.NMR* **19** (2001) 199-208.
48. N. Bonander, J. Leckner, H. Guo, G. Karlsson and L. Sjölin. *Eur.J.Biochem.* **267** (2000) 4511-4519.
49. E.I. Solomon, M.J. Baldwin and M.D. Lowery. *Chem.Rev.* **92** (1992) 521-542.
50. F.E. Dodd, Z.H.L. Abraham, R.R. Eady and S. Samar Hasnain. *Acta Cryst.* **D56** (2000) 690-696.
51. K. Cheung, R.W. Strange and S. Samar Hasnain. *Acta Cryst.* **D56** (2000) 697-704.
52. A. Ciocchetti, A.R. Bizzarri and S. Cannistraro. *Biophys.Chem.* **69** (1997) 185-198.
53. M.A. Halcrow. *Angew.Chem.Int.Ed.* **40** (2001) 346-349.
54. D.L. Tennent and D.R. McMillin. *J.Am.Chem.Soc.* **101** (1979) 2307-2311.
55. B.G. Malmström. *Eur.J.Biochem.* **223** (1994) 711-718.
56. A. Donaire, B. Jiménez, J.-M. Moratal, J.F. Hall and S. Samar Hasnain. *Biochem.* **40** (2001) 837-846.
57. E.I. Solomon, K.W. Penfield, A.A. Gewirth, M.D. Lowery, S.E. Shadle, J.A. Guckert and L.B. LaCroix. *Inorg.Chim.Acta* **243** (1996) 67-78.
58. E.I. Solomon and M.D. Lowery. *Science* **259** (1993) 1575-1581.
59. H. Nar, R. Huber, A. Messerschmidt, A.C. Filippou, M. Barth, M. Jaquinod, M. van de Kamp and G.W. Canters. *Eur.J.Biochem.* **205** (1992) 1123-1129.
60. D. Bashford, M. Karplus and G.W. Canters. *J.Mol.Biol.* **203** (1988) 507-510.
61. I.M.C. van Amsterdam, M. Ubbink, M. van den Bosch, F. Rotsaert, J. Sanders-Loehr and G.W. Canters. (2002) *in preparation*.
62. T.J. Thamann, P. Frank, L.J. Willis and T.M. Loehr. *Proc.Natl.Acad.Sci.USA* **79** (1982) 6396-6400.
63. G. Battistuzzi, M. Borsari, G.W. Canters, E.C. de Waal, L. Loschi, G. Warmerdam and M. Sola. *Biochem.* **40** (2001) 6707-6712.
64. J.J.R. Fraústo do Silva and R.J.P. Williams. *The Biological Chemistry of the Elements. The Inorganic Chemistry of Life*, Paperback ed., Oxford Univ. Press, 1991.
65. H.B. Gray and E.I. Solomon. "Electronic Structures of Blue Copper Centers in Proteins". In *Copper Proteins*, Spiro, T. G., Ed., Wiley, 1981.
66. W.E. Blumberg. In *The Biochemistry of Copper*, Peisach, J., Aisen, P., Blumberg, W. E., Eds., Academic, 1966, 49-65.
67. A.S. Brill and G.F. Bryce. *J.Chem.Phys.* **48** (1968) 4398-4404.
68. R.J.P. Williams. *Inorg.Chim.Acta Rev.* **5** (1971) 137-155.
69. S. Larsson, A. Broo and L. Sjölin. *J.Phys.Chem.* **99** (1995) 4860-4865.
70. A.A. Gewirth and E.I. Solomon. *J.Am.Chem.Soc.* **110** (1988) 3811-3819.
71. M. van Gastel, J.W.A. Coremans, H. Sommerdijk, M.C. van Hemert and E.J.J. Groenen. *J.Am.Chem.Soc.* **124** (2002) 2035-2041.
72. E.T. Adman, G.W. Canters, H.A.O. Hill and N.A. Kitchen. *FEBS Lett.* **143** (1982) 287-292.
73. N. Bonander, T. Vänngård, L.-C. Tsai, V. Langer, H. Nar and L. Sjölin. *Prot.Struct.Func.Gen.* **27** (1997) 385-394.
74. J.W.A. Coremans, O.G. Poluektov, E.J.J. Groenen, G.W. Canters, H. Nar and A. Messerschmidt. *J.Am.Chem.Soc.* **116** (1994) 3097-3101.
75. T. den Blaauwen and G.W. Canters. *J.Am.Chem.Soc.* **115** (1993) 1121-1129.
76. C.M. Groeneveld and G.W. Canters. *Eur.J.Biochem.* **153** (1985) 559-564.
77. B.G. Karlsson, L.C. Tsai, H. Nar, J. Sanders-Loehr, N. Bonander, V. Langer and L. Sjolin. *Biochem.* **36** (1997) 4089-4095.
78. J.-M. Moratal, J. Salgado, A. Donaire, H.R. Jiménez and J. Castells. *J.Chem.Soc.Chem.Comm.* (1993) 110-112.
79. H. Nar, A. Messerschmidt, R. Huber, M. van de Kamp and G.W. Canters. *FEBS Lett.* **306** (1992) 119-124.

80. M. Piccioli, C. Luchinat, T.J. Mizoguchi, B.E. Ramirez, H.B. Gray and J.H. Richards. *Inorg.Chem.* **34** (1995) 737-742.
81. I. Pozdnyakova, J. Guidry and P. Wittung-Stafshede. *J.Biol.Inorg.Chem.* **6** (2001) 182-188.
82. R.W. Strange, L.M. Murphy, B.G. Karlsson, B. Reinhammar and S. Samar Hasnain. *Biochem.* **35** (1996) 16391-16398.
83. M.A. Webb, C.N. Kiser, J.H. Richards, A.J. Di Bilio, H.B. Gray, J.R. Winkler and G.R. Loppnow. *J.Phys.Chem. B* **104** (2000) 10915-10920.
84. C.W.G. Hoitink and G.W. Canters. *J.Biol.Chem.* **267** (1992) 13836-13842.
85. T. Koslowski. *Phys.Chem.Chem.Phys.* **3** (2001) 1497-1503.
86. L.B. LaCroix, D.W. Randall, A.M. Nersissian, C.W.G. Hoitink, G.W. Canters, J.S. Valentine and E.I. Solomon. *J.Am.Chem.Soc.* **120** (1998) 9621-9631.
87. N. Bonander, B.G. Karlsson and T. Vanngard. *Biochim.Biophys.Acta* **1251** (1995) 48-54.
88. R. Guzzi, A. Stirpe, M.Ph. Verbeet and L. Sportelli. *Eur.Biophys.J.Biophys.Lett.* **30** (2001) 171-178.
89. B. Rizzuti, L. Sportelli and R. Guzzi. *Biophys.Chem.* **94** (2001) 107-120.
90. G. Battistuzzi, M. Borsari, L. Loschi, M.C. Menziani, F. de Rienzo and M. Sola. *Biochem.* **40** (2001) 6422-6430.
91. G.W. Canters and G. Gilardi. *FEBS Lett.* **325** (1993) 39-48.
92. G.W. Canters, U. Kolczak, F. Armstrong, L.J.C. Jeuken, R. Camba and M. Sola. *Farad.Disc.* (2000) 205-220.
93. S.J. Kroes, J. Salgado, G. Parigi, C. Luchinat and G.W. Canters. *J.Biol.Inorg.Chem.* **1** (1996) 551-559.
94. G.W. Canters and G. van Pouderoyen. *Biosens.Bioelec.* **9** (1994) 637-645.
95. M. van de Kamp, R. Floris, F.C. Hali and G.W. Canters. *J.Am.Chem.Soc.* **112** (1990) 907-908.
96. P.J. Hart, A.M. Nersissian, R.G. Herrmann, R.M. Nalbandyan, J.S. Valentine and D. Eisenberg. *Prot.Science* **5** (1996) 2175-2183.
97. C. Arcangeli, A.R. Bizzarri and S. Cannistraro. *Biophys.Chem.* **92** (2001) 183-199.
98. A.R. Bizzarri and S. Cannistraro. *Chem.Phys.Lett.* **349** (2001) 503-510.
99. I. Chandrasekhar, G.M. Clore, A. Szabo, A.M. Gronenhorn and B.R. Brooks. *J.Mol.Biol.* **226** (1992) 239-250.
100. V. Lounnas, S.K. Lüdemann and R.C. Wade. *Biophys.Chem.* **78** (1999) 157-182.
101. S. Melchionna, M. Falconi and A. Desideri. *J.Chem.Phys.* **108** (1998) 6033-6041.
102. L.W. Ungar, N.F. Scherer and G.A. Voth. *Biophys.J.* **72** (1997) 5-17.
103. J.O.A. de Kerpel, K. Pierloot, U. Ryde and B.O. Roos. *J.Phys.Chem. B* **102** (1998) 4638-4647.
104. J.O.A. de Kerpel, K. Pierloot and U. Ryde. *J.Phys.Chem. B* **103** (1999) 8375-8382.
105. K.W. Penfield, A.A. Gewirth and E.I. Solomon. *J.Am.Chem.Soc.* **107** (1985) 4519-4529.
106. L.B. LaCroix, S.E. Shadle, Y.N. Wang, B.A. Averill, B. Hedman, K.O. Hodgson and E.I. Solomon. *J.Am.Chem.Soc.* **118** (1996) 7755-7768.
107. M.D. Lowery and E.I. Solomon. *Inorg.Chim.Acta* **200** (1992) 233-243.
108. R.J.P. Williams. *J.Mol.Catal.* **30** (1985) 1-26.
109. R.J.P. Williams. *Eur.J.Biochem.* **234** (1995) 363-381.
110. P.W. Atkins. *Physical Chemistry*, Oxford University Press, 1994.
111. A. Szabo and N.S. Ostlund. *Modern Quantum Chemistry (Revised First Edition)*, McGraw-Hill, 1989.
112. R. McWeeny. *Methods of Molecular Quantum Mechanics*, Academic Press, 1989.
113. J.M. Seminario and P. Politzer. *Modern Density Functional Theory : A Tool for Chemistry*, Elsevier, 1995.
114. P. Hohenberg and W. Kohn. *Phys.Rev.* **136** (1964) B864-B871.
115. E. Bright Wilson. "DFT explained". In *Structural Chemistry and Biology*, 1968, 753-760.
116. W. Kohn and L.J. Sham. *Phys.Rev.* **140** (1965) A1133-A1138.
117. G. te Velde, F.M. Bickelhaupt, E.J. Baerends, C. Fonseca Guerra, S.J.A. van Gisbergen, J.G. Snijders and T. Ziegler. *J.Comput.Chem.* **22** (2001) 931-967.
118. C. Fonseca Guerra, J.G. Snijders, G. te Velde and E.J. Baerends. *Theor.Chem.Acc.* **99** (1998) 391-403.
119. S.H. Vosko, L. Wilk and M. Nusair. *Can.J.Phys.* **58** (1980) 1200-1211.
120. A.D. Becke. *Phys.Rev. A* **38** (1988) 3098-3100.
121. J.P. Perdew. *Phys.Rev. B* **33** (1986) 8822-8824. Erratum: *Ibid.* **34** (1986) 7406.
122. A.D. Becke. *J.Chem.Phys.* **98** (1993) 5648-5652.
123. A.D. Becke. *J.Chem.Phys.* **107** (1997) 8554-8560.
124. J.P. Perdew, S. Kurth, A. Zupan and P. Blaha. *Phys.Rev.Lett.* **82** (1999) 2544-2547.

125. J.P. Perdew, S. Kurth, A. Zupan and P. Blaha. *Phys.Rev.Lett.* **82** (1999) 5179-5179.
126. W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell and P.A. Kollman. *J.Am.Chem.Soc.* **117** (1995) 5179-5197.
127. W.F. van Gunsteren and H.J.C. Berendsen. *GROningen MOlecular Simulation (GROMOS) Library Manual*, Biomos, 1996.
128. A.K. Rappé, C.J. Casewit, K.S. Colwell, W. Goddard III and W.M. Skiff. *J.Am.Chem.Soc.* **114** (1992) 10024-10035.
129. N.L. Allinger, X. Zhou and J. Bergsma. *J.Mol.Str. (THEOCHEM)* **312** (1994) 69-83.
130. P.M. Morse. *Phys.Rev.* **34** (1929) 57-64.
131. A.A. Frost and B. Musulin. *J.Am.Chem.Soc.* **76** (1954) 2045-2048.
132. A.A. Frost and B. Musulin. *J.Chem.Phys.* **22** (1954) 1017-1020.
133. A.H. de Vries, P.Th. van Duijnen, A.H. Juffer, J.A.C. Rullmann, J.P. Dijkman, H. Merenga and B.T. Thole. *J.Comput.Chem.* **16** (1995) 37-55. Erratum: *Ibid.* (1995) 1445-1446.
134. A.H. de Vries, P.Th. van Duijnen, R.W.J. Zijlstra and M. Swart. *J.El.Spectr.Rel.Phen.* **86** (1997) 49-56.
135. P.Th. van Duijnen, A.H. Juffer and J.P. Dijkman. *J.Mol.Str. (THEOCHEM)* **260** (1992) 195-205.
136. P.Th. van Duijnen. "Embedding in Quantum Chemistry: the Direct Reaction Field Approach". In *New Challenges in Computional Quantum Chemistry*, Broer, R., Aerts, P. J. C., Bagus, P. S., Eds., Dept. of Chemical Physics and Material Science, 1994.
137. P.Th. van Duijnen and A.H. de Vries. *Int.J.Quant.Chem., QCS* **29** (1995) 523-532.
138. P.Th. van Duijnen and A.H. de Vries. *Int.J.Quant.Chem.* **60** (1996) 1111-1132.
139. P.Th. van Duijnen and M. Swart. *J.Phys.Chem. A* **102** (1998) 2399-2407.
140. P.Th. van Duijnen, M. Swart and F. Grozema. "QM/MM Calculation of (Hyper-)Polarizabilities with the DRF Approach". In *ACS Symposium Series 712*, ACS, 1998, 220-232.
141. P.Th. van Duijnen, F.C. Grozema and M. Swart. *J.Mol.Str. (THEOCHEM)* **464** (1999) 193-200.
142. R.W.J. Zijlstra, F.C. Grozema, M. Swart, B.L. Feringa and P.Th. van Duijnen. *J.Phys.Chem. A* **105** (2001) 3583-3590.
143. A.H. Juffer, E.F.F. Botta, B.A.M. van Keulen, A. van der Ploeg and H.J.C. Berendsen. *J.Comp.Phys.* **97** (1991) 144-171.
144. A.H. Juffer. PhD Thesis: *On the Modelling of Solvent Mean Force Potentials—From Liquid Argon to Solvated Macromolecules.* Rijks*universiteit* Groningen, The Netherlands, 1993.
145. A.H. Juffer and H.J.C. Berendsen. *Mol.Phys.* **79** (1993) 623-644.
146. H.L. Friedman. *Mol.Phys.* **29** (1975) 1533-1543.
147. W. Rocchia, E. Alexov and B. Honig. *J.Phys.Chem. B* **105** (2001) 6507-6514.
148. L. Jensen, J.G. Snijders and P.Th. van Duijnen. "PhD Study: Coupling DRF with DFT".
149. *Combined Quantum Mechanical and Molecular Mechanical Methods, ACS Symposium Series 712*, J. Gao and M.A. Thompson, Eds., 1998.
150. U. Röthlisberger. *ACS Symp.Ser.* **712** (1998) 264-274.
151. S. Dapprich, I. Komaromi, K.S. Byun, K. Morokuma and M.J. Frisch. *J.Mol.Str. (THEOCHEM)* **461-462** (1999) 1-21.
152. M. Svensson, S. Humbel, R.D.J. Froese, T. Matsubara, S. Sieber and K. Morokuma. *J.Phys.Chem.* **100** (1996) 19357-19363.
153. T. Vreven and K. Morokuma. *J.Chem.Phys.* **113** (2000) 2969-2975.
154. J. Sauer and M. Sierka. *J.Comput.Chem.* **21** (2000) 1470-1493.
155. T.K. Woo, L. Cavallo and T. Ziegler. *Theor.Chem.Acc.* **100** (1998) 307-313.
156. T.K. Woo, P.M. Margl, L. Deng, L. Cavallo and T. Ziegler. *Catal.Today* **50** (1999) 479-500.
157. T.K. Woo, P.E. Blöchl and T. Ziegler. *J.Mol.Str. (THEOCHEM)* **506** (2000) 313-334.
158. M. Swart. *Int.J.Quant.Chem.* (2002) *accepted.*
159. W.F. van Gunsteren and H.J.C. Berendsen. *Angew.Chem.Int.Ed.* **29** (1990) 992-1023.
160. M.P. Allen and D.J. Tildesley. *Computer Simulation of Liquids*, Clarendon Press, 1987.
161. D. Frenkel and B. Smit. *Understanding Molecular Simulation : from Algorithms to Applications*, Academic Press, 1996.
162. S. Toxvaerd. *Mol.Phys.* **72** (1991) 159-168.
163. W.G. Hoover. *Phys.Rev. A* **31** (1985) 1695-1697.

164. W.G. Hoover. *Phys.Rev. A* **34** (1986) 2499-2500.
165. S. Nosé. *J.Chem.Phys.* **81** (1984) 511-519.
166. H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. Dinola and J.R. Haak. *J.Chem.Phys.* **81** (1984) 3684-3690.
167. A.J. Stone and M. Alderton. *Mol.Phys.* **56** (1985) 1047-1064.
168. A.J. Stone. *Chem.Phys.Lett.* **83** (1981) 233-239.
169. S. Bachrach. "Population Analysis and Electron Density from Quantum Mechanics". In *Reviews in Computational Chemistry Vol. 5*, Lipkowitz, K. B., Boyd, D. B., Eds., VCH, 1994.
170. K.B. Wiberg and P.R. Rablen. *J.Comput.Chem.* **14** (1993) 1504-1518.
171. E. Sigfridsson and U. Ryde. *J.Comput.Chem.* **19** (1998) 377-395.
172. l.E. Chirlian and M.M. Francl. *J.Comput.Chem.* **8** (1987) 894-905.
173. C.M. Breneman and K.B. Wiberg. *J.Comput.Chem.* **11** (1990) 361-373.
174. B.T. Thole and P.Th. van Duijnen. *Theor.Chim.Acta* **63** (1983) 209-221.
175. A.J. Stone. *The Theory of Intermolecular Forces, Appendix B*, Clarendon Press, 1996.
176. E. Clementi and G. Corongiu. *Methods and Techniques in Computational Chemistry METECC-95*, STEF, 1995.
177. ADF 2.3.3, Theoretische Chemie, Vrije Universiteit, Amsterdam, 1997.
178. ADF 1999.03, SCM, Amsterdam, 1999.
179. A.D. Buckingham. "Basic Theory of Intermolecular Forces: Applications to Small Molecules". In *Intermolecular Interactions: From Diatomics to Biopolymers*, Pullman, B., Ed., John Wiley & Sons, 1978, 1-67.
180. A.D. Buckingham. *Quart.Rev.Chem.Soc., Lond.* **13** (1959) 183-214.
181. A.D. Buckingham, R.L. Disch and D.A. Dunmur. *J.Am.Chem.Soc.* **90** (1968) 3104-3107.
182. F. de Proft, F. Tielens and P. Geerlings. *J.Mol.Str. (THEOCHEM)* **506** (2000) 1-8.
183. M. Swart, P.Th. van Duijnen and J.G. Snijders. *J.Comput.Chem.* **22** (2001) 79-88.
184. J.M. Seminario. *Int.J.Quant.Chem., QCS* **30** (1996) 59-65.
185. Y.P. Varshni. *Rev.Mod.Phys.* **29** (1957) 664-682.
186. E.B. Wilson Jr., J.C. Decius and P.C. Cross. *Molecular Vibrations—The Theory of Infrared and Raman Vibrational Spectra*, McGraw-Hill, 1955.
187. ADF 2000.02, SCM, Amsterdam, 2000.
188. T. Helgaker, J. Gauss, P. Jørgensen and J. Olsen. *J.Chem.Phys.* **106** (1997) 6430-6440.
189. J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh and C. Fiolhais. *Phys.Rev. B* **46** (1992) 6671-6687. Erratum: *Ibid.* **48** (1993) 4978.
190. G. Herzberg. *Molecular Spectra and Molecular Structure. I. Spectra of Diatomic Molecules*, Van Nostrand Company, 1950.
191. A. Blondel and M. Karplus. *J.Comput.Chem.* **17** (1996) 1132-1141.
192. A. Warshel and M. Levitt. *J.Mol.Biol.* **103** (1976) 227-249.
193. J. Gao, D. Habibollazadeh and L. Shao. *J.Phys.Chem.* **99** (1995) 16460-16467.
194. M.A. Thompson. *J.Phys.Chem.* **100** (1996) 14492-14507.
195. A.H. de Vries and P.Th. van Duijnen. *Int.J.Quant.Chem.* **57** (1996) 1067-1076.
196. F.C. Grozema and P.Th. van Duijnen. *J.Phys.Chem. A* **102** (1998) 7984-7989.
197. R.W.J. Zijlstra, P.Th. van Duijnen and A.H. de Vries. *Chem.Phys.* **204** (1996) 439-446.
198. J. Cioslowski, J. Sauer, J. Hetzenegger, T. Karcher and T. Hierstetter. *J.Am.Chem.Soc.* **115** (1993) 1353-1359.
199. T.R. Furlani and J. Gao. *J.Org.Chem.* **61** (1996) 5492-5497.
200. S.J. Stuart and B.J. Berne. *J.Phys.Chem.* **100** (1996) 11934-11943.
201. B.D. Bursulaya, D.A. Zichi and H.J. Kim. *J.Phys.Chem.* **99** (1995) 10069-10074.
202. A.H. de Vries. PhD Thesis: *Modelling Condensed-Phase Systems. From Quantum Chemistry to Molecular Models.* Rijks*universiteit* Groningen, The Netherlands, 1995.
203. T.H. Dunning Jr. and P.J. Hay. "Gaussian Basis Sets for Molecular Calculations". In *Methods of Electronic Structure Theory*, Schaefer III, H. F., Ed., 1977, 1-27.
204. M. Dupuis, A. Farazdel, S.P. Karma and S.A. Maluendes. "HONDO: a General Atomic and Molecular Electronic Structure System". In *MOTECC-90*, Clementi, E., Ed., ESCOM, 1990, 277-342.
205. H.-J. Werner and W. Meyer. *Mol.Phys.* **31** (1976) 855-872.

206. A.J. Sadlej. *Theor.Chim.Acta* **79** (1991) 123-140.
207. A.J. Sadlej and M. Urban. *J.Mol.Str. (THEOCHEM)* **234** (1991) 147-171.
208. A.J. Sadlej. *Theor.Chim.Acta* **81** (1991) 45-63.
209. A.J. Sadlej. *Theor.Chim.Acta* **81** (1991) 339-354.
210. S.J.A. van Gisbergen, V.P. Osinga, O.V. Gritsenko, R. van Leeuwen, J.G. Snijders and E.J. Baerends. *J.Chem.Phys.* **105** (1996) 3142-3151.
211. S.J.A. van Gisbergen, J.G. Snijders and E.J. Baerends. *J.Chem.Phys.* **103** (1995) 9347-9354.
212. S.J.A. van Gisbergen, J.G. Snijders and E.J. Baerends. *Chem.Phys.Lett.* **259** (1996) 599-604.
213. E.J. Baerends, D.E. Ellis and P. Ros. *Chem.Phys.* **2** (1973) 41-51.
214. C. Fonseca Guerra, O. Visser, J.G. Snijders, G. te Velde and E.J. Baerends. "Parallelisation of the Amsterdam Density Functional Programme". In *Methods and Techniques for Computational Chemistry, METECC-5*, Clementi, E., Corongiu, G., Eds., STEF, 1995, 303-395.
215. G. te Velde and E.J. Baerends. *J.Comp.Phys.* **99** (1992) 84-98.
216. R. van Leeuwen and E.J. Baerends. *Phys.Rev. A* **49** (1994) 2421-2431.
217. B.T. Thole. *Chem.Phys.* **59** (1981) 341-350.
218. J.J.P. Stewart. *MOPAC 93.00 Manual*, Fujitsu Ltd., Tokyo, 1993.
219. J.J.P. Stewart. MOPAC93, Fujitsu Ltd., Tokyo, 1993.
220. C.J.F. Böttcher and P. Bordewijk. *Theory of Electric Polarization Vol. II*, 2nd ed., Elsevier, 1978.
221. J. Applequist, J.R. Carl and J.K. Fung. *J.Am.Chem.Soc.* **94** (1972) 2952-2960.
222. T.H. Dunning Jr. *J.Chem.Phys.* **90** (1989) 1007-1023.
223. H. Koch, P. Jørgensen and T. Helgaker. *J.Chem.Phys.* **104** (1996) 9528-9530.
224. E. van Lenthe, E.J. Baerends and J.G. Snijders. *J.Chem.Phys.* **99** (1993) 4597-4610.
225. M. Dupuis, A. Marquez, E.R. Davidson, H. Nakano, K. Hirao, J.P. Flament, A.M. Chaka, M. Aida, J. Li, G.D. Hawkins, D.A. Liotard, C.J. Cramer and D.G. Truhlar. HONDO98, 1998.
226. E.I. Proynov, S. Sirois and D.R. Salahub. *Int.J.Quant.Chem.* **64** (1997) 427-446.
227. C. Lee, W. Yang and R.G. Parr. *Phys.Rev. B* **37** (1988) 785-789.
228. J.P. Perdew, K. Burke and M. Ernzerhof. *Phys.Rev.Lett.* **77** (1997) 3865-3868. Erratum: *Ibid.* **78** (1997) 1396.
229. J.P. Perdew. In *Electronic Structure of Solids 1991*, Ziesche, P., Eschrig, H., Eds., Akademie, 1991, 11-20.
230. Y. Zhang and W. Yang. *Phys.Rev.Lett.* **80** (1998) 890.
231. B. Hammer, L.B. Hansen and J.K. Norskov. *Phys.Rev. B* **59** (1999) 7413-7421.
232. W. Klopper and H.P. Lüthi. *Chem.Phys.Lett.* **262** (1996) 546-552.
233. K. Pierloot, B.J. Persson and B.O. Roos. *J.Phys.Chem.* **99** (1995) 3465-3472.
234. N. Matsuzawa, J. Seto and D.A. Dixon. *J.Phys.Chem. A* **101** (1997) 9391-9398.
235. R.S. Mulliken. *J.Chem.Phys.* **23** (1955) 1833-1840.
236. F.L. Hirshfeld. *Theor.Chim.Acta* **44** (1977) 129-138.
237. S.H. Pine. *Organic Chemistry*, Fifth ed., McGraw-Hill, 1987.
238. M. Kitamura, S. Okada, S. Suga and R. Noyori. *J.Am.Chem.Soc.* **111** (1989) 4028-4036.
239. M. Kitamura, S. Suga, M. Niwa, R. Noyori, Z.X. Zhai and H. Suga. *J.Phys.Chem.* **98** (1994) 12776-12781.
240. M. Kitamura, S. Suga, M. Niwa and R. Noyori. *J.Am.Chem.Soc.* **117** (1995) 4832-4842.
241. M. Kitamura, S. Suga, H. Oka and R. Noyori. *J.Am.Chem.Soc.* **120** (1998) 9800-9809.
242. M. Kitamura, H. Oka and R. Noyori. *Tetrahedr.* **55** (1999) 3605-3614.
243. M. Yamakawa and R. Noyori. *J.Am.Chem.Soc.* **117** (1995) 6327-6335.
244. M. Yamakawa and R. Noyori. *Org.Metal.* **18** (1999) 128-133.
245. D. Guillaneux, S.H. Zhao, O. Samuel, D. Rainford and H.B. Kagan. *J.Am.Chem.Soc.* **116** (1994) 9430-9439.
246. R.M. Kellogg and R.P. Hof. *J.Chem.Soc.Perkin Trans. 1* (1996) 1651-1657.
247. R.M. Kellogg. *Phosph.Sulf.Silic.* **120-121** (1997) 61-76.
248. M.A. Poelert. PhD Thesis: *Stereoselective Transformations of Ephedra Alkaloids, Chapter 7.* Rijks*universiteit* Groningen, The Netherlands, 1993.
249. R.P. Hof, M.A. Poelert, N. Peper and R.M. Kellogg. *Tetrahedr.Asym.* **5** (1994) 31-34.
250. K. Fitzpatrick, R. Hulst and R.M. Kellogg. *Tetrahedr.Asym.* **6** (1995) 1861-1864.
251. J. Kang, D.S. Kim and J.I. Kim. *Synlett* (1994) 842-844.
252. J. Kang, J.W. Lee and J.I. Kim. *J.Chem.Soc.Chem.Comm.* (1994) 2009-2010.

253. J. Kang, J.W. Kim, J.W. Lee, D.S. Kim and J.I. Kim. *Bull.Kor.Chem.Soc.* **17** (1996) 1135-1142.
254. J. Kang, J.B. Kim, J.W. Kim and D. Lee. *J.Chem.Soc.Per.Trans. 2* (1997) 189-194.
255. J. Kang, J.B. Kim, J. Kim and D. Lee. *Bull.Kor.Chem.Soc.* **19** (1998) 475-481.
256. W.J. Hehre, L. Radom, P.R. Schleyer and J.A. Pople. "Chapter 6.3.9: Thermodynamic Properties". In *Ab Initio Molecular Orbital Theory*, Wiley, 1986.
257. T. Oka and F.J. Simpson. *Bioch.Bioph.Res.Comm.* **43** (1971) 1-5.
258. H.K. Hund, J. Breuer, F. Lingens, J. Huttermann, R. Kappl and S. Fetzner. *Eur.J.Biochem.* **263** (1999) 871-878.
259. F. Fusetti, K.H. Schroter, R.A. Steiner, P.I. van Noort, T. Pijning, H.J. Rozeboom, K.H. Kalk, M.R. Egmond and B.W. Dijkstra. *Structure* **10** (2002) 259-268.
260. I.M. Kooter, R.A. Steiner, B.W. Dijkstra, P.I. van Noort, M.R. Egmond and M. Huber. *Eur.J.Biochem.* (2002) *accepted.*
261. R.A. Steiner. PhD Thesis: *Mechanistic Studies on the Copper-Dependent Quercetin 2,3-Dioxygenase from A. japonicus.* Rijks*universiteit* Groningen, The Netherlands, 2002.
262. R.A. Steiner, K.H. Kalk and B.W. Dijkstra. (2002) *in preparation.*
263. L. Barhacs, J. Kaizer and G. Speier. *J.Org.Chem.* **65** (2000) 3449-3452.
264. L. Barhacs, J. Kaizer, J. Pap and G. Speier. *Inorg.Chim.Acta* **320** (2001) 83-91.
265. A.H. de Vries, P.Th. van Duijnen and A.H. Juffer. *Int.J.Quant.Chem., QCS* **27** (1993) 451-466.
266. F.C. Grozema, R.W.J. Zijlstra and P.Th. van Duijnen. *Chem.Phys.* **246** (1999) 217-227.
267. R.W.J. Zijlstra, P.Th. van Duijnen, B.L. Feringa, T. Steffen, K. Duppen and D.A. Wiersma. *J.Phys.Chem. A* **101** (1997) 9828-9836.
268. T. Funabiki and T. Yamazaki. *J.Mol.Catal.* **150** (1999) 37-47.
269. C.G. Nordstrom. *Suomen Kem.* **B41** (1968) 351-353.
270. A. Nishinaga, T. Tojo, H. Tomita and T. Matsuura. *J.Chem.Soc.Perkin Trans. 1* (1979) 2511-2516.
271. M. Schulz and K. Kirschke. In *Organic Peroxides Vol. III*, Swern, D., Ed., Wiley, 1972, 2nd ed., 67-140.
272. E. Balogh-Hergovich, J. Kaizer and G. Speier. *Inorg.Chim.Acta* **256** (1997) 9-14.
273. L. Barhacs, J. Kaizer and G. Speier. *J.Mol.Catal.* **172** (2001) 117-125.
274. E. Balogh-Hergovich, J. Kaizer, G. Speier, G. Huttner and L. Zsolnai. *Inorg.Chim.Acta* **304** (2000) 72-77.
275. A.W. Addison, H.M.J. Hendriks, J. Reedijk and L.K. Thompson. *Inorg.Chem.* **20** (1981) 103-110.
276. F. Jiang, K.D. Karlin and J. Peisach. *Inorg.Chem.* **32** (1993) 2576-2582.
277. E.T. Adman and L.H. Jensen. *Isr.J.Chem.* **21** (1981) 8-12.
278. E.T. Adman, R.E. Stenkamp, L.C. Sieker and L.H. Jensen. *J.Mol.Biol.* **12** (1978) 35-47.
279. H.J.C. Berendsen, D. van der Spoel and R. van Drunen. *Comp.Phys.Comm.* **95** (1995) 43-56.
280. A. Messerschmidt and G.W. Canters, *personal communication.*
281. H.J.C. Berendsen, J.R. Grigera and T.P. Straatsma. *J.Phys.Chem.* **91** (1987) 6269-6271.
282. W. Weltner. *Magnetic Atoms and Molecules*, Dover, 1989.
283. A. Abragam and B. Bleaney. *Electron Paramagnetic Resonance of Transition Ions*, Dover, 1986.
284. E. van Lenthe, P.E.S. Wormer and A. van der Avoird. *J.Chem.Phys.* **107** (1997) 2488-2498.
285. O.L. Malkina, J. Vaara, B. Schimmelpfennig, M. Munzarová, V.G. Malkin and M. Kaupp. *J.Am.Chem.Soc.* **122** (2000) 9206-9218.
286. E. van Lenthe, A. van der Avoird, W.R. Hagen and E.J. Reijerse. *J.Phys.Chem. A* **104** (2000) 2070-2077.
287. S. Patchkovskii and T. Ziegler. *J.Chem.Phys.* **111** (1999) 5730-5740.
288. M. Stein, E. van Lenthe, E.J. Baerends and W. Lubitz. *J.Am.Chem.Soc.* **123** (2001) 5839-5840.
289. M. Stein, E. van Lenthe, E.J. Baerends and W. Lubitz. *J.Phys.Chem. A* **105** (2001) 416-425.
290. S. Patchkovskii and T. Ziegler. *J.Am.Chem.Soc.* **122** (2000) 3506-3516.
291. S. Patchkovskii and T. Ziegler. *J.Phys.Chem. A* **105** (2001) 5490-5497.
292. E. van Lenthe, A. van der Avoird and P.E.S. Wormer. *J.Chem.Phys.* **108** (1998) 4783-4796.
293. G. Schreckenbach and T. Ziegler. *J.Phys.Chem. A* **101** (1997) 3388-3399.
294. A.A. Gewirth, S.L. Cohen, H.J. Schugar and E.I. Solomon. *Inorg.Chem.* **26** (1987) 1133-1146.
295. P.H. Kasai, E.B. Whipple and W. Weltner Jr. *J.Chem.Phys.* **44** (1966) 2581-2591.
296. T. Ichikawa, H. Yoshida and L. Kevan. *J.Chem.Phys.* **75** (1981) 2485-2488.
297. M. Sano, K. Taniguchi and H. Yamatera. *Chem.Lett.* (1980) 1285-1286.
298. M. Munzarová and M. Kaupp. *J.Phys.Chem. A* **103** (1999) 9966-9983.

299. P.R.T. Schipper, O.V. Gritsenko, S.J.A. van Gisbergen and E.J. Baerends. *J.Chem.Phys.* **112** (2000) 1344-1352.
300. M. Gruning, O.V. Gritsenko, S.J.A. van Gisbergen and E.J. Baerends. *J.Chem.Phys.* **114** (2001) 652-660.
301. S.V. Didziulis, S.L. Cohen, A.A. Gewirth and E.I. Solomon. *J.Am.Chem.Soc.* **110** (1988) 250-268.
302. J.W.A. Coremans, M. van Gastel, O.G. Poluektov, E.J.J. Groenen, T. den Blaauwen, G. van Pouderoyen, G.W. Canters, H. Nar, C. Hammann and A. Messerschmidt. *Chem.Phys.Lett.* **235** (1995) 202-210.
303. J.W.A. Coremans, O.G. Poluektov, E.J.J. Groenen, G.W. Canters, H. Nar and A. Messerschmidt. *J.Am.Chem.Soc.* **118** (1996) 12141-12153.
304. J.W.A. Coremans, O.G. Poluektov, E.J.J. Groenen, G.W. Canters, H. Nar and A. Messerschmidt. *J.Am.Chem.Soc.* **119** (1997) 4726-4731.
305. W.P. Anderson, W.D. Edwards and M.C. Zerner. *Inorg.Chem.* **25** (1986) 2728-2732.
306. M.C. Zerner, G.H. Loew, R.F. Kirchner and U.T. Mueller-Westerhoff. *J.Am.Chem.Soc.* **102** (1980) 589-599.
307. D.W. Clack and M.S. Farrimond. *J.Chem.Soc. A* (1971) 299-304.
308. G.C. Allen and D.W. Clack. *J.Chem.Soc. A* (1970) 2668-2672.
309. F.C. Grozema, L.P. Candeias, M. Swart, P.Th. van Duijnen, J. Wildeman, G. Hadziioanou, L.D.A. Siebbeles and J.M. Warman. *J.Chem.Phys.* (2002) *submitted.*
310. X. Assfeld, N. Ferré and J.-L. Rivail. *ACS Symp.Ser.* **712** (1998) 234-249.
311. D. Vanderbilt. *Phys.Rev. B* **41** (1990) 7892-7895.
312. D.A. Estrin, C. Tsoo and S.J. Singer. *Chem.Phys.Lett.* **184** (1984) 571-578.
313. A. Bergner, M. Dolg, W. Küchle, H. Stoll and H. Preuß. *Mol.Phys.* **80** (1993) 1431-1441.
314. U. Chandra Singh and P.A. Kollman. *J.Comput.Chem.* **7** (1986) 718-730.
315. I. Antes and W. Thiel. *ACS Symp.Ser.* **712** (1998) 50-65.
316. R.J. Hall, S.A. Hindle, N.A. Burton and I.H. Hillier. *J.Comput.Chem.* **21** (2000) 1433-1441.
317. M. Swart, J.G. Snijders. *Theor.Chem.Acc.* (2002) *to be submitted.*
318. M. Swart. *pdb2adf*, Groningen, 2002.
319. L.J.C. Jeuken, M. Ubbink, J.H. Bitter, P. van Vliet, W. Meyer-Klaucke and G.W. Canters. *J.Mol.Biol.* **299** (2000) 737-755.
320. A. Donaire, J. Salgado and J.-M. Moratal. *Biochem.* **37** (1998) 8659-8673.
321. J.-M. Moratal, J. Salgado, A. Donaire, H.R. Jimenez and J. Castells. *Inorg.Chem.* **32** (1993) 3587-3588.
322. A.J. Dibilio, T.K. Chang, B.G. Malmstrom, H.B. Gray, B.G. Karlsson, M. Nordling, T. Pascher and L.G. Lundberg. *Inorg.Chim.Acta* **200** (1992) 145-148.
323. S.L. Dong, J.A. Ybe, M.H. Hecht and T.G. Spiro. *Biochem.* **38** (1999) 3379-3385.
324. L.J.C. Jeuken, P. van Vliet, M.Ph. Verbeet, R. Camba, J.P. McEvoy, F.K. Armstrong and G.W. Canters. *J.Am.Chem.Soc.* **122** (2000) 12186-12194.
325. O. Farver, L.J.C. Jeuken, G.W. Canters and I. Pecht. *Eur.J.Biochem.* **267** (2000) 3123-3129.
326. M. Milani, L. Andolfi, S. Cannistraro, M.Ph. Verbeet and M. Bolognesi. *Acta Cryst.* **D57** (2001) 1735-1738.
327. I.M.C. van Amsterdam, M. Ubbink, L.J.C. Jeuken, M.Ph. Verbeet, O. Einsle, A. Messerschmidt and G.W. Canters. *Chem.Eur.J.* **7** (2001) 2398-2406.
328. J.-M. Moratal, J. Salgado, A. Donaire, H.R. Jimenez, J. Castells and M.J. Martinez Ferrer. *Magn.Res.Chem.* **31** (1993) S41-S46.
329. C. Dennison, T. Kohzuma, W. McFarlane, S. Suzuki and A.G. Sykes. *Inorg.Chem.* **33** (1994) 3299-3305.
330. J.-M. Moratal, A. Romero, J. Salgado, A. Perales-Alarcón and H.R. Jiménez. *Eur.J.Biochem.* **228** (1995) 653-657.

SPECIAL

S.1 *Samenvatting*

Bij scheikundig onderzoek wordt al snel gedacht aan laboratoria met allerlei pruttelende kolfjes en reageerbuisjes, vervuiling door chemische stoffen en dergelijke stereotypes. Een reëel beeld van de werkelijkheid is het echter niet helemaal. Weliswaar wordt een groot deel van het huidige onderzoek nog gedaan op voorgenoemde experimentele wijze, maar een "nieuwe" trend binnen chemisch onderzoek is de toepassing van schone chemie, namelijk met behulp van berekeningen.

Heel erg nieuw is deze manier van onderzoek doen niet, de beginselen (van de zogenaamde kwantumchemie) werden immers al in het begin van de 20ste eeuw geformuleerd. Lange tijd echter was de toepassing van kwantumchemie beperkt tot relatief kleine moleculen, maar door de steeds krachtiger en sneller wordende computers is het in de laatste jaren ook mogelijk om grote (bio-) systemen aan te pakken. Vooral de beschrijving van metaaleiwitten met behulp van berekeningen is de laatste jaren enorm in opmars.

Onderzoek naar metaaleiwitten is van belang aangezien ze verschillende elementaire processen regelen binnen het menselijk lichaam, alsmede in het planten en dierenrijk. Door de toepassing van metaalatomen in eiwitten kunnen deze processen uitgevoerd worden onder "lichaams-eigen" omstandigheden. Kopereiwitten hebben binnen de complexe chemie in onder andere het menselijk lichaam vaak de rol van elektronenoverdrachtseiwitten; dat wil zeggen dat op de ene plek een elektron wordt opgenomen, waarna het eiwit naar een andere plek gaat en daar het elektron weer afgeeft. Het gebruik van koper hiervoor is "natuurlijk" aangezien het zowel in de mono-valente vorm (Cu^{I}) als in de di-valente vorm (Cu^{II}) een sterke binding heeft met het eiwit, wat gepaard gaat met een grote stabiliteit. Daarnaast is het vaak zo dat het eiwit het koperatoom in een bepaalde conformatie dwingt, die het midden houdt tussen de mono- en di-valente voorkeursconformatie. Dit vergemakkelijkt de elektronenoverdracht waardoor het eiwit zijn functie nog efficiënter kan uitvoeren.

De toepassing van kwantumchemie op metaaleiwitten vertoont dezelfde opmars als het gebruik van DFT (*Density Functional Theory*) binnen chemisch onderzoek. Tot een aantal jaren geleden waren voorbeelden van beide relatief gezien schaars, maar door de ontwikkelingen op het gebied van DFT (zoals de ontwikkeling van betere [GGA] potentialen, efficiëntere programmatuur onder andere door parallel computing technieken) is het tegenwoordig bijna standaard om DFT berekeningen te gebruiken in onderzoek.

Toch blijven biosystemen als (metaal-)eiwitten te groot om volledig te worden beschreven met kwantumchemie. Daarnaast wordt er een gedetailleerde methode gebruikt voor grote delen van het systeem waar een minder nauwkeurige beschrijving (met behulp van speciaal ontwikkelde *force fields*) meer dan voldoende is om de relevante processen goed te kunnen beschrijven. Voor de dynamische eigenschappen blijken deze *force fields* vaak een zeer goede beschrijving te geven, tenzij er een metaalatoom aanwezig is in het eiwit. De interacties van een metaalatoom met het eiwit blijken namelijk moeilijk te generaliseren in een algemeen toepasbaar *force field*. De ontwikkeling van een *force field* specifiek voor een bepaald (type) eiwit was een van de doelen van het onderzoek beschreven in dit proefschrift.

Kopereiwitten worden ingedeeld in diverse types, naar gelang hun structurele en spectroscopische eigenschappen. Type 2 kopereiwitten vertonen electronparamagnetische spectra die gelijk zijn aan die van "normale" koper-complexen, type 3 kopereiwitten vertonen helemaal geen paramagnetische activiteit, terwijl type 1 kopereiwitten gekarakteriseerd worden door een felblauwe kleur. Daarom

worden ze in de regel *blauwe koper*eiwitten genoemd. Het eiwit dat werd bestudeerd in dit proefschrift is azurine, een type 1 kopereiwit, dat waarschijnlijk elektronenoverdracht als functie heeft en dat bestaat uit 128 (*Pseudomonas aeruginosa*) of 129 (*Alcaligenes denitrificans*) aminozuren. Net als alle andere type 1 kopereiwitten is het koperatoom in het actieve centrum gebonden aan drie sterke liganden (twee histidine en een cysteine aminozuurzijketens) die ongeveer in één vlak liggen met het koperatoom. Daarnaast zijn er twee axiale aminozuurresiduen (glycine en methionine), die gecoördineerd zijn met het koperatoom. Het glycineresidu is speciaal in dit opzicht aangezien azurine het enige type 1 kopereiwit is waar dit residu coördineert met de koper.

Een aantal methodes werd ontwikkeld om de benodigde *force field* parameters voor deze kopereiwitten te halen uit DFT gegevens. De eerste methode (MDC ladingsanalyse) levert atomaire ladingen op die representatief zijn voor het bestudeerde systeem: de meest relevante moleculaire en atomaire eigenschappen (multipolen) voor de bepaling van de Coulomb potentiaal worden *exact* gerepresenteerd. De tweede methode (IntraFF) haalt uit de berekende (*Hessian*) matrix met de tweede afgeleiden van de energie naar de atomaire coördinaten, waardes voor krachtsconstanten die gebruikt kunnen worden in *force field* berekeningen cq. simulaties voor interacties waarvoor geen standaardwaarde beschikbaar is.

Als eerste werden echter een aantal eigenschappen van het gebruik van DFT vastgesteld, met name de nauwkeurigheid van de voorspelde waardes van moleculaire polariseerbaarheden, geometrieën van moleculen, alsmede een vergelijking van de nieuw ontwikkelde ladingsanalyse met andere methodes. Uit deze resultaten blijkt dat DFT zeer nauwkeurig de geometrie of polariseerbaarheid van of ladingsverdeling binnen een molecuul kan voorspellen, met een nauwkeurigheid en efficiëntie die moeilijk te bereiken is met andere kwantumchemische methodes.

Daarnaast werd DFT gebruikt om het voorgestelde mechanisme van twee reacties te toetsen. De eerste bestudeerde reactie is afkomstig uit de organische chemie, waarbij een zinkatoom betrokken is. De verschillende invloeden van een aminoalcohol of een aminothiol op de asymmetrische additie van dialkylzinkmoleculen aan aldehyde werden onderzocht, waarbij de enantiomere selectiviteit alsmede de versnellende invloed van de katalysator verklaard kunnen worden met behulp van de berekende resultaten. De tweede reactie gaat al meer richting het einddoel van het promotie-onderzoek, aangezien daarin een enzym werd bestudeerd dat een koperatoom bevat in het actieve centrum. De verschillende stadia die leiden van de complexatie van het substraat naar productvorming werden bestudeerd, waarbij een aantal belangrijke intermediairen gelokaliseerd werden.

Hiermee komen we dan bij de studies aan kopereiwitten, waarbij het azurine eiwit bekeken werd in zijn natuurlijke (*wildtype*) vorm, alsmede in een aantal andere vormen waarbij er verschillende aminozuren vervangen zijn door andere. De invloed van deze vervangingen (*mutaties*) op relevante eigenschappen van het eiwit is experimenteel uitgebreid getest, wat een uitstekend referentiekader biedt om de berekeningen aan te toetsen. De ladingsverdeling in het actieve centrum van azurine (zowel *wildtype* als een aantal mutaties ervan) blijkt een coherente structuur te hebben: de manier waarop de elektronen verdeeld zijn binnen het actieve centrum varieert nauwelijks als men kijkt naar de verschillende azurinemoleculen; in alle gevallen is de totale lading op koper een stuk lager dan zijn formele lading, terwijl er een grote mate van delokalisatie is over de drie histidine en cysteine liganden, alsmede over de axiale methionine groep. Dit suggereert een functie als ligand ook voor methionine, terwijl glycine beter omschreven kan worden als coördinerende groep.

Naast het bestuderen van de ladingsverdeling van het actieve centrum werden ook krachtsconstanten bepaald voor de bindingen tussen koper en de vijf aminozuren in het actieve centrum. De waardes hiervoor blijken sterk afhankelijk te zijn van de gebruikte geometrie van het centrum, wat een direct gevolg is van het anharmonische verloop van de binding. Aangezien de

geometrie direct overgenomen is van de experimenteel bepaalde kristalstructuur met bijbehorende onnauwkeurigheid van 0.1-0.2 Å, moeten de harmonische *force field* parameters zorgvuldig bekeken worden. Beter zou het dan zijn om de anharmonische parameters, die ook geleverd worden door de IntraFF methode, te gebruiken. Evenwel, een moleculaire dynamica simulatie aan *wildtype* azurine met de harmonische IntraFF parameters en de MDC ladingen levert een stabiel actief centrum en eiwit op in de simulatie, met koper-ligand afstanden die overeenkomen met experimentele gegevens en vibratiefrequenties voor de koper-ligandbindingen die in het juiste bereik van 200-500 cm^{-1} liggen.

Metaaleiwitten worden vaak gekarakteriseerd met behulp van UV/VIS en EPR/ESR spectroscopie. De spectra van de eiwitten kunnen daarom dienen als referentiemateriaal om de berekende karakteristieken mee te vergelijken. In het geval van EPR-spectroscopie gaat het dan om de zogenaamde g-tensor, die bepaald wordt door een ongepaard elektron. Het gebruik van DFT om deze tensor te berekenen blijkt voor systemen met metaalatomen soms problematisch, met een onderschatting van de z-waarde van de tensor tot 50%. Ook de voorspelling van de hyperfijnkoppelingstensor van koper is soms matig, met z-waardes die tot drie keer te groot zijn. Daarentegen worden de hyperfijnkoppelingstensoren van andere atomen vaak redelijk goed beschreven, zeker in het geval van de kopereiwitten waar de verschillen tussen de twee histidine residuen in het actieve centrum goed voorspeld worden.

De UV/VIS spectra (of excitatie-energieën) van kopereiwitten kunnen met DFT alleen berekend worden voor de gereduceerde toestand (met koper formeel in zijn +1 redox toestand), en niet voor de geoxideerde toestand (formele redox toestand +2). Daarom werden de UV/VIS spectra voorspeld met semi-empirische CNDO/INDO berekeningen, zowel voor het actieve centrum alleen als het actieve centrum in de omgeving van het eiwit en oplosmiddel. Hierbij bleek de aanwezigheid van de omgeving een beslissende rol te spelen voor de berekende excitatie-energieën. De parameters voor koper in de gereduceerde toestand werden hierbij verkregen door de berekende semi-empirische excitatie-energieën te fitten aan de berekende DFT energieën.

Om niet afhankelijk te zijn van de kristalstructuur met de bijbehorende onnauwkeurigheid van 0.1-0.2 Å, en om de invloed van de eiwitomgeving op de geometrie van het actieve centrum te bepalen, werden QM/MM berekeningen uitgevoerd. In deze berekeningen werd het hele eiwit (inclusief een schil van watermoleculen rondom) meegenomen in de berekening, waarbij het actieve centrum werd beschouwd met DFT (QM systeem) en de rest van het eiwit (en oplosmiddelmoleculen) met een klassiek *force field* (MM systeem), dat gespecialiseerd is voor het beschrijven van biosystemen zoals eiwitten. Een nieuw model (AddRemove) voor de directe koppeling van de QM en MM systemen werd daarvoor ontwikkeld, waarin in de DFT berekening waterstofatomen worden toegevoegd aan het actieve centrum om de valentie van het QM systeem te waarborgen. Achteraf worden de interacties van de toegevoegde waterstoffen met de echte QM atomen gecorrigeerd, zodat in principe de *artificiele* toevoeging van de waterstoffen geen effect heeft op de geometrie en/of energie.

De geometrie van een aantal azurinemoleculen (*wildtype*, gemuteerd en metaalgesubstitueerd) werd met QM/MM geoptimaliseerd, waarbij het aantal atomen varieerde van ca. 2200 tot 14.000, afhankelijk van het aantal watermoleculen dat meegenomen werd in de berekening. De geoptimaliseerde geometrie van het actieve centrum vertoont over het algemeen een grote gelijkenis met experimenteel bepaalde structuren (ofwel kristalstructuur of EXAFS). De enige echte discrepantie treedt op voor de gereduceerde toestand van de Met121Gln azurine-mutant, die in berekeningen vrijwel identiek is aan de gereduceerde *wildtype* azurinestructuur, maar die qua kristalstructuur een behoorlijke vervorming vertoont. De afwijking is dermate groot, en voor de rest de overeenkomst tussen berekende en experimentele structuur ook, dat het niet te verwachten is dat er iets mis is met de berekeningen. Het lijkt er eerder op dat er een ander (onverwacht) effect een rol speelt.

S.2

Summary

Chemical research is often associated with laboratories with all kinds of boiling pots and test-tubes, toxic waste by chemical compounds and more of these stereotypes. However, it is not a true image of reality. Even though a large portion of current chemical research is still carried out in the aforementioned experimental manner, a "new" trend within chemical research is the application of clean chemistry, by performing theoretical calculations.

This way of doing research is not exactly new, as the basic principles (of the so-called quantum chemistry) had already been formulated in the beginning of the 20th century. However, for a long time the application of quantum chemistry was limited to relatively small molecules. With the computers getting faster and more powerful, it has been possible in the last couple of years to also take a look at large (bio-)systems. Especially the investigation of metalloproteins by quantum chemical calculations has increased considerably.

The study of metalloproteins is of major importance as they regulate several elementary processes within the human body, as well as in plants and animal life. The application of metal atoms in proteins enable that these processes be carried out under normal "human" circumstances. Copper proteins often function as electron transfer proteins within the complex chemistry in amongst others the human body; that is, at one place an electron is taken up, after which the protein moves to another place and delivers the electron. Using copper for this is "natural" as it is both in the mono-valent (Cu^{I}) as the divalent (Cu^{II}) form strongly bound to the protein, which is accompanied by a large stability of the protein. Furthermore, often the protein forces the copper atom in a certain conformation, which is midway between the mono- and divalent preferred conformations. This enhances electron transfer and enables the protein to perform its function even more efficiently.

The application of quantum chemistry to metalloproteins shows the same increase as the usage of DFT (*Density Functional Theory*) within chemical research. Until a few years ago examples of both were relatively scarce, but developments within the field of DFT (as the construction of improved [GGA] potentials, more efficient computer programs with the help of for instance parallel computing techniques) have enabled a situation where it is nowadays almost standard to incorporate (DFT) calculations in (bio)chemical studies.

Still, biosystems like (metallo)proteins remain to be too large to be described completely by quantum chemistry. Moreover, a detailed method would be used for large parts of the system where a less accurate description (by using specially designed *force fields*) would suffice to describe the relevant processes well. These standard *force fields* are known to give a good description for capturing dynamical properties, unless there is a metal atom present in the protein, as the interactions of the metal atom with the protein are difficult to generalize in a generally applicable *force field*. The development of a *force field* specific for a certain (type of) metalloprotein was one of the aims of the research described in this thesis.

Copper proteins are classified into several types, depending on their structural and spectroscopic properties. Type 2 copper proteins show electron paramagnetic spectra similar to "normal" copper complexes, type 3 proteins show no paramagnetic activity at all, while type 1 copper proteins are characterized by a bright blue color. Therefore, the latter are usually referred to as blue copper proteins. The protein studied in this thesis is azurin, a type 1 copper protein that probably serves as electron transfer protein and that consists of 128 (*Pseudomonas aeruginosa*) or 129 (*Alcaligenes*

denitrificans) amino acid residues. The copper atom is, just like in all other type 1 copper proteins, in the active site bound to three strong ligands (two histidine and one cysteine amino acid residues) that lie approximately in one plane with the copper atom. There are also two axial amino acid residues (glycine and methionine) that are coordinated towards the copper. In this respect, the glycine residue is special as azurin is the only type 1 copper protein where this residue is coordinated towards copper.

Several methods were developed to obtain the necessary *force field* parameters for the copper proteins from DFT data. The first method (MDC charge analysis) gives atomic charges that are representative for the studied system; the most relevant molecular and atomic properties (multipoles) for the evaluation of the Coulomb potential are represented *exactly*. The second method (IntraFF) extracts from the computed Hessian matrix, a matrix with the second derivatives of the energy with respect to the atomic coordinates, values for force constants that can be used in *force field* calculations or simulations for interactions for which no standard force constant is available.

At first however, a few properties of the usage of DFT were determined, in particular the accuracy of predicted values of molecular polarizabilities, geometries of molecules, and a comparison of the newly developed charge analysis with other methods. The results show that DFT can predict the geometry or polarizability of a molecule or the charge distribution within a molecule very accurately, with an accuracy and efficiency that is difficult to attain with other quantum chemical methods.

DFT was also used to check the proposed mechanism of two reactions. The first reaction that was studied is coming from organic chemistry and involves a zinc atom. The different effects of an aminoalcohol or an aminothiol on the asymmetric addition of dialkylzinc molecules to an aldehyde were investigated; the enantiomeric selectivity as well as the effect on the reaction rate of the catalyst can be explained by the computed results. The second reaction is more in the direction of the aim of the PhD research, as an enzyme was investigated with a copper atom present in the active site. The different stages that lead from the initial complexation of the substrate to the formation of the product were studied, where a few important intermediates were localized.

We thenarrive at the studies on copper proteins, where the azurin protein was investigated in its natural (*wildtype*) form as well as in several other forms where different amino acid residues have been replaced by others. The influence of these replacements (*mutations*) on relevant properties of the protein has been extensively tested experimentally, which gives a solid framework against which the calculations can be checked. The charge distribution within the active site of azurin (both *wildtype* as mutated) is shown to have a coherent structure: the way in which the electrons are spread over the site varies little when looking at the different azurin molecules; in all cases the total charge on copper is considerably lower than its formal value, with a large amount of delocalization of charge over the three in-plane ligands, as well as the axial methionine group. This suggests a role as ligand also for the methionine, while the other axial group, glycine, can be better described as coordinating group.

Apart from studying the charge distribution in the active site also force constant values were determined for the bonds between copper and the five amino acid residues in the active site. These values were found to be strongly dependent on the geometry of the site that was used, which is a direct consequence of the anharmonicity of the bond. As the geometry used was taken directly from the experimentally determined crystal structure with a corresponding uncertainty of 0.1-0.2 Å, the harmonic *force field* parameters should be judged cautiously. It might be better to use the anharmonic parameters, which are also given by the IntraFF method. On the other hand, a molecular dynamics simulation of *wildtype* azurin using the harmonic IntraFF parameters and the MDC charges results in a stable active site and protein in the simulation, with copper-ligand distances that compare well with experimental values and vibrational frequencies that are in the right range of 200-500 cm^{-1}.

Metalloproteins are often characterized by UV/VIS and EPR/ESR spectroscopy. The spectra of the proteins can therefore be used as reference for comparison with the computed characteristics. In the case of EPR spectroscopy, the so-called g-tensor is dealt with, which is determined by an unpaired electron. Using DFT to compute this tensor seems to be sometimes problematic for systems containing metal atoms, with an underestimation of the z-value of the tensor by up to 50 %. Also the prediction of the hyperfine coupling tensor of copper is sometimes mediocre, with z-values up to three times too large. On the contrary, the hyperfine couplings of the other atoms are reasonably well predicted, especially in the case of copper proteins where the differences between the two histidine residues in the active site are well predicted.

The UV/VIS spectra (or excitation energies) of copper proteins can be computed with DFT only for the reduced state (with copper formally in its +1 redox state), not for the oxidized state (formal redox state +2). Therefore, the UV/VIS spectra were predicted by semi-empirical CNDO/INDO calculations, both for the active site alone as well as the site surrounded by the protein and solvent. The presence of the surroundings was shown to play a decisive role for the computed excitation energies. In this study, the parameters for copper in the reduced state were obtained from fitting the computed semi-empirical excitation energies to the computed DFT energies.

In order to avoid the dependency on the crystal structure with the corresponding uncertainty of 0.1-0.2 Å as well as to investigate the influence of the protein environment on the geometry of the active site, QM/MM calculations were performed. In these calculations the complete protein (including a shell of water molecules around it) was taken into account in the calculation, where the active site was treated with DFT (QM system) and the rest of the protein (and solvent molecules) with a classical *force field* (MM system), which is designed for the description of biosystems like proteins. A new model (AddRemove) for the direct coupling of the QM and MM systems was developed, in which hydrogen atoms are added in the DFT calculation to satisfy the valences of the QM system. Afterwards the interactions of the added hydrogens with the real QM atoms are corrected for; therefore the *artificial* introduction of the hydrogens has in principle no effect on the geometry and/or energy.

The geometry of a number of azurin molecules (*wildtype*, mutated and metal substituted) was optimized by the QM/MM method, where the total number of atoms varied from ca. 2200 to 14.000, depending on the number of water molecules that were taken into consideration in the calculation. The optimized geometry of the active site is in general in good agreement with experimentally observed structures (either crystal or EXAFS). The only real discrepancy is found for the reduced state of the Met121Gln mutant; the computed structure is similar to the reduced *wildtype* azurin structure, while the crystal structure shows a large deformation. The difference between the computed and experimental structure is that large, and contradictory to the general agreement for the other molecules, that it is not expected that there is something wrong with the calculations. It rather seems that some (unexpected) other effect plays a decisive role here.

S.3 *Dankwoord*

Mijn dank gaat allereerst uit naar Jaap Snijders. Jaap, ik kan me nog goed herinneren hoe ik de eerste keer bij je kwam om te solliciteren voor een AIO-plaats over *Solid State Physics* (waar Freddie Kootstra vorig jaar november op gepromoveerd is). Het eerste wat je zei was dat je een andere onderzoekplaats voor mij in gedachten had. Ik moest wel een aantal maanden wachten voor ik daarmee kon beginnen, maar aangezien het onderzoek perfect aansloot bij mijn hoofdvak onderzoek, had ik daar geen problemen mee. Een ander iets dat ik me goed kan herinneren is de schaterende bulderlach die soms door de gang klonk en waaruit direct bleek: “ah, Jaap is ook aanwezig”. Ook de gezellige momenten tijdens de groepsuitstapjes en de jaarlijkse barbecue bij jullie thuis komen dan weer in beeld. Ook heb ik veel van je geleerd over het beheren van een cluster van computers, zowel met de IBM’s als de Pentium’s.

Mijn beide andere promotores, Herman Berendsen en Gerard Canters, wil ik voornamelijk bedanken voor de discussies tijdens de bijeenkomsten die we af en toe hadden. Herman, ik kan me goed herinneren dat je tijdens die bijeenkomsten vaak optrad als bemiddelaar tussen de theoretische en experimentele wereld, waarin je vaak duidelijkheid bracht over wat er nu eigenlijk bedoeld werd; uiteindelijk bleek vaak dat hetzelfde doel voor ogen stond, alleen was er soms wat spraakverwarring. Jouw rol tijdens die bijeenkomsten en de discussies die we los daarvan hebben gehad, heb ik zeer gewaardeerd. Ik was ook zeer blij en aangenaam verrast over je zeer gedetailleerde commentaar bij het corrigeren van mijn proefschrift; ik heb er veel van geleerd en zal er in de toekomst zeker profijt van hebben. Hetzelfde geldt voor Gerard, ook jou wil ik van harte bedanken voor het kritisch doorlezen van mijn proefschrift. Je opmerkingen waren van enorm belang om het bioanorganische aspect van het onderzoek duidelijk te maken, en daarbij vooral om aan een theoreet duidelijk te maken waar een experimentalist in geïnteresseerd is. Je gastvrijheid en het vrijmaken van tijd als ik weer eens een vraag of probleem had, heb ik zeer gewaardeerd. Ik voelde me altijd op mijn gemak, of het nu was op het Gorlaeus, een conferentie of tijdens een bijeenkomst.

Marieke, ik weet niet wat ik zonder jou had gemoeten; de simulaties die je gedaan hebt, zijn van grote waarde geweest voor mijn onderzoek, zowel om aan te geven wat nu eigenlijk de beperkingen van de standaard force fields zijn als om mijn berekende waardes in de praktijk te toetsen. Ook was het nooit een probleem om weer 3 uur in de trein naar Leiden te zitten, aangezien ik wist dat er altijd wel weer iets interessants te bespreken of te doen was.

Alan, I would like to thank you very much for your invaluable help on getting the simulations working. Your approach of handling scientific issues is something we have benefited greatly from in this project. I would like to thank you also for being member of the reading committee, and taking the time to struggle through this somewhat lengthy thesis.

Ook de beide andere leden van de leescommissie, Bauke en Koop, wil ik van harte bedanken voor het plaatsnemen in de commissie en het doorlezen van het manuscript.

Edgar Groenen en Martina Huber wil ik van harte bedanken voor de menige discussies over EPR spectroscopie aan, en g-tensoren van, kopereiwitten. Edgar, je hebt niet alleen duidelijk gemaakt wat EPR/ENDOR/ESEEM nu eigenlijk inhoudt en wat je er mee kan bereiken, maar ook dat berekeningen aan g-tensoren geen trivialiteit zijn. Jouw manier van uitleggen van zowel theorie als experiment maakt dat het lijkt alsof het simpel is, terwijl het dat in feite absoluut niet is. Als er één ding is wat ik geleerd heb van de studie beschreven in hoofdstuk 7, is het dat wel.

Daarnaast zijn er verschillende mensen in de Theoretische Chemie groep die ik van harte wil bedanken om uiteenlopende redenen. Allereerst natuurlijk Johan. Jouw nuchterheid is van onschatbare waarde gebleken voor de stabiliteit van de computerclusters, en dus ook voor mijn onderzoek. Ook je vele nuttige tips met betrekking tot allerlei details over het gebruik van de Mac, het upgraden naar nieuwe software en het systeembeheer van de clusters hebben zeker invloed gehad op het totstandkomen van dit boekje. Freddie, mijn buddy vanaf het eerste uur, wil ik bedanken voor de vele nachten die we samen doorbrachten in den vreemde, en ook de goede tijd die we hadden als kamergenoten in de eerste jaren van mijn promotietijd. De vier jaar zijn daardoor een stuk sneller voorbijgegaan, met af en toe een leuke ontspanning zoals bijvoorbeeld het organiseren van het Tourspel. Tommie, mijn roommate van de laatste tijd, jou wil ik van harte bedanken voor de vooral rustige tijd die we samen doorbrachten, ieders luisterend naar zijn eigen muziek en vooral communicerend via ICknoei. Een paar dingen die me wel stoorden waren dat je telkens meer bureaus in beslag nam (nu ik weg ben, geloof ik zelfs alle vijf) en dat je Nederlands telkens beter en beter werd, zodat snel praten (zoals Freddie en ik in het begin nog wel eens deden) ook al niet meer hielp. Rosie, tja, waar moet ik beginnen. Zowel op wetenschappelijk als op culinair als op sociaal niveau heb ik een geweldige tijd met je gehad, en ik denk (en hoop) dat dat in de toekomst zo zal blijven, en ik denk dat ik hier mede namens Elvira spreek. Liviu, I'll always remember you as a good colleague, with a great sense of humour and like Isa would say, a very funny laugh. Lasse, het was hoog tijd dat er iemand het werk ging doen waar jij nu mee bezig bent, en ik denk dat jij er de juiste persoon voor bent. Paul, ik hoop dat we elkaar nog vaak ontmoeten in de toekomst, want het is altijd een plezier om je ontmoeten en te spreken, of het nu gaat over eten, een film, de keiharde wetenschap of gewoon lekker de laatste roddels. Robert, Meta, Arjen, Alexandrina, de tijd samen in de groep was relatief gezien kort, maar zeer zeker aangenaam. Nienke, het was leuk om je als kamergenote te hebben. Coen y Carmen, ik vond het jammer dat jullie al zo snel naar España gingen, en vond het leuk om jullie in Barcelona weer te ontmoeten. Renu, Kechen, it was a pleasure having you around in our group. Wim, jouw verhalen zijn altijd interessant om aan te horen, en je zat er ook nooit om één verlegen. Ria, jouw positieve invloed op de groep wordt denk ik soms wel eens onderschat, maar ik heb veel van je geleerd door de jaren heen. Last but not least at all, Piet. Ik denk dat je wel weet wat voor een bewondering ik heb voor je wetenschappelijke instelling en je soms heerlijk kritische attitude. Ondanks dat mijn promotie-onderzoek weinig te maken had met het DRF model, ben ik blij dat we toch nog zoveel hebben kunnen samenwerken en dat uiteindelijk het INDRF programma ook nog van pas kwam voor mijn eigen onderzoek. Ik hoop dat deze samenwerking ook in de toekomst nog door mag gaan.

The cooperation with Roberto Steiner and Bruno Rizutti is greatly acknowledged. Roberto, working with you on *quercetinase* posed a real challenge on the use of ADF, which made it a lot of fun. Bruno, *grazie mille* for your help on getting the azurin force field right.

De ondersteuning vanuit Amsterdam, indien nodig, van Stan en Olivier is iets waarvoor ik niet genoeg dank kan zeggen. Alle nuttige suggesties bij bug-fixes, upgraden naar een nieuwe versie, CVS problemen en submitten naar de trunk, lijken vanzelfsprekend, maar zijn zeer waardevol gebleken. Ook de vele discussies over het ADF programma en het uitbreiden ervan heb ik als zeer stimulerend ervaren, wat zeer zeker heeft bijgedragen aan de inhoud van dit boekje.

There are many people I would like to thank for creating a very pleasant atmosphere at summer schools and conferences: Vincent, Bernd, Aaron, Isa, Per-Erik, Andrea, Llano, Jesus, Elena, Spiridoula, Perdita. I think I will miss the games of Uno and Asshole very much.

Alle andere mensen die ik nog vergeten ben, maar die op een of andere manier hebben bijgedragen aan het tot stand komen van dit boekje of het aangenaam maken van de jaren in Groningen, wil ik van harte bedanken.

Mijn paranimfen, Rob Zijlstra en Ferdinand Grozema, wil ik bedanken voor de vele nuttige discussies over van alles en nog wat, of het nou ging om het DRF model, de zinkcomplexen, het Tourspel of een ander vrijdagmiddag projectje. Rob, ik denk dat je, wetenschappelijk gezien, een grotere invloed op me hebt gehad dan je zelf had gedacht. Het formuleren van de eigenlijke doelstelling van een onderzoek is iets waarbij je een grote invloed op me hebt gehad, maar ook dat het hebben van plezier net zo belangrijk is als het bereiken van een wetenschappelijk resultaat. Groxe, jou wil ik bedanken voor je heerlijk nuchtere Groningse opvattingen, of het nu gaat om scheikunde, een lekker biertje of wat dan ook. Het was altijd een verademing om je weer aan de telefoon te hebben, mede omdat je soms kritische opmerkingen vaak wel hout sneden.

Blijft er nog over mijn derde paranimf, die eigenlijk geen paranimf mag zijn, maar het zo graag wilde zijn. Elvira, zonder jou had ik deze laatste jaren nooit kunnen volhouden of aangekund. Je voortdurende ondersteunende woorden waren belangrijke steunpilaren in moeilijke tijden, en de goede tijden waren nooit zo leuk geweest zonder jou. Het was daarom altijd een plezier om naar huis te gaan, ook al betekende dat soms (vaak) dat ik je opzadelde met een lastig probleem waar ik mee zat. Ik weet zeker dat ik het zonder jou niet gered had.

Papa en mama, ondanks dat jullie vandaag niet meer mee mochten maken, wil ik jullie bedanken voor alle steun die jullie me in het verleden hebben gegeven om door te gaan, en hoop dat jullie een beetje trots op me zijn.

S.4 *List of publications*

M. Swart, M. van den Bosch, H.J.C. Berendsen, G.W. Canters and J.G. Snijders
"Active Site Geometries of Copper Proteins Obtained with a QM/MM Approach"
in preparation

F.C. Grozema, L. P. Candeias, M. Swart, P.Th. van Duijnen, J. Wildeman, G. Hadziioanou, L.D.A. Siebbeles and J.M. Warman
"Theoretical and Experimental Studies of the Opto-Electronic Properties of Positively Charged Oligo(Phenylene Vinylene)s: Effects of Chain Length and Alkoxy Substitution"
J.Chem.Phys. (2002) *submitted*

R.A. Steiner, M. Swart, J.G. Snijders and B.W. Dijkstra
"Density Functional Theory Investigation of the Reaction Mechanism of *A. japonicus* 2,3QD"
in preparation

M. Swart, M. van den Bosch, H.J.C. Berendsen, G.W. Canters and J.G. Snijders
"Calculations on the Spectral Characteristics of Copper Proteins"
in preparation

M. Swart
"AddRemove: a New Link Model for Use in QM/MM Studies"
Int.J.Quant.Chem. (2002) *accepted*

L. Jensen, M. Swart, P.Th. van Duijnen and J.G. Snijders
"Medium Perturbations on the Molecular Polarizability Calculated within a Localized Dipole Interaction Model"
J.Chem.Phys. (2002) *submitted*

M. Remko, P.Th. van Duijnen and M. Swart
"Theoretical Study of Molecular Structure, Tautomerism, and Geometrical Isomerism of N-Methyl and N-Phenyl Substituted Cyclic Imidazolines, Oxazolines and Thiazolines"
Phys.Chem.Chem.Phys. (2002) *in press*

P.Th. van Duijnen, A.H. de Vries, M. Swart and F.C. Grozema
"Polarizabilities in the Condensed Phase and the Local Fields Problem. A Direct Reaction Field Formulation"
J.Chem.Phys. (2002) *submitted*

M. Swart, M. van den Bosch, H.J.C. Berendsen, G.W. Canters and J.G. Snijders
"IntraFF: Obtaining Reliable Force Constants from Quantum Chemical Calculations"
to be submitted

M. Swart, M. van den Bosch, H.J.C. Berendsen, G.W. Canters, A.E. Mark and J.G. Snijders
"Modelling of the Redox Potential of Copper Proteins: Density Functional Theory and Molecular Dynamics Results for *Wildtype* and Mutated Azurin"
to be submitted

M. Swart, R.W.J. Zijlstra, F.C. Grozema, J.G. Snijders and R.M. Kellogg
"Asymmetric Addition of Dialkylzincs to Aldehydes Studied by High-Level Ab Initio Calculations. Similarities and Differences between Aminoalcohol and Aminothiol Promoted Reactions"
to be submitted

M. Swart and P.Th. van Duijnen
"DRF90: a Polarizable Force Field"
J.Comput.Chem. (2002) *submitted*

T. la Cour Jansen, M. Swart, L. Jensen, P.Th. van Duijnen, J.G. Snijders and K. Duppen
"Collision Effects in the Nonlinear Raman Response of Liquid Carbon Disulfide"
J.Chem.Phys. **116** (2002) 3277-3285

R.W.J. Zijlstra, F.C. Grozema, M. Swart, B.L. Feringa and P.Th. van Duijnen
"Solvent Induced Charge Separation in the Excited States of Symmetrical Ethylene: a Direct Reaction Field Study"
J.Phys.Chem. A **105** (2001) 3583-3590

M. Swart, P.Th. van Duijnen and J.G. Snijders
"A Charge Analysis Derived from an Atomic Multipole Expansion"
J.Comput.Chem. **22** (2001) 79-88

F.C. Grozema, R.W.J. Zijlstra, M. Swart and P.Th. van Duijnen
"On the Iodine-Benzene Charge-Transfer Complex: Potential Energy Surface and Transition Probabilities Studied at several Levels of Theory"
Int.J.Quant.Chem. **75** (1999) 709-723

P.Th. van Duijnen, F.C. Grozema and M. Swart
"Some Applications of the Direct Reaction Field Approach"
J.Mol.Str. (THEOCHEM) **464** (1999) 193-200

P.Th. van Duijnen, M. Swart and F.C. Grozema
"QM/MM Calculation of (Hyper-)Polarizabilities with the Direct Reaction Field Approach"
ACS Symposium Series **712** (1999) 220-232

M. Swart, P.Th. van Duijnen and J.G. Snijders
"Mean Polarizabilities of Organic Molecules: a Comparison of Restricted Hartree Fock, Density Functional Theory and Direct Reaction Field Results"
J.Mol.Str. (THEOCHEM) **458** (1999) 11-17

P.Th. van Duijnen and M. Swart
"Molecular and Atomic Polarizabilities: Thole's Model Revisited"
J.Phys.Chem. A **102** (1998) 2399-2407

A.H. de Vries, P.Th. van Duijnen, R.W.J. Zijlstra and M. Swart
"Thole's Interacting Polarizability Model in Computational Chemistry Practice"
J.El.Spectr.Rel.Phen. **86** (1997) 49-56