

chapter one

INTRODUCTION

*and
historical
background*



CONTENTS

1.1	<i>Computational chemistry</i>	3
1.2	<i>Copper proteins</i>	4-6
1.3	<i>Overview thesis</i>	7-8

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 chemistry¹, 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 nature²⁻⁶³. A substantial part of these investigations involves proteins where copper is incorporated, since it is one of the most abundant metals found in proteins⁶⁴. 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 ($5500 \text{ M}^{-1} \text{ cm}^{-1}$) 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 ($\sim 0.0060 \text{ cm}^{-1}$) than that found for simple copper complexes ($\sim 0.0160 \text{ cm}^{-1}$). 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 closely phenylalanine residue.

As indicated above, type 1 copper proteins exhibit an intense blue color due to a strong absorption band at $\sim 625 \text{ nm}$ with a molar absorption coefficient of $5500 \text{ M}^{-1} \text{ cm}^{-1}$ (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 $d-d$ transition⁶⁵⁻⁶⁷, while Williams proposed it to be a ligand to metal charge transfer band^{65,68}. Later studies substantiated the ligand to metal transition⁶⁹⁻⁷¹, 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 $d-d$ 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).

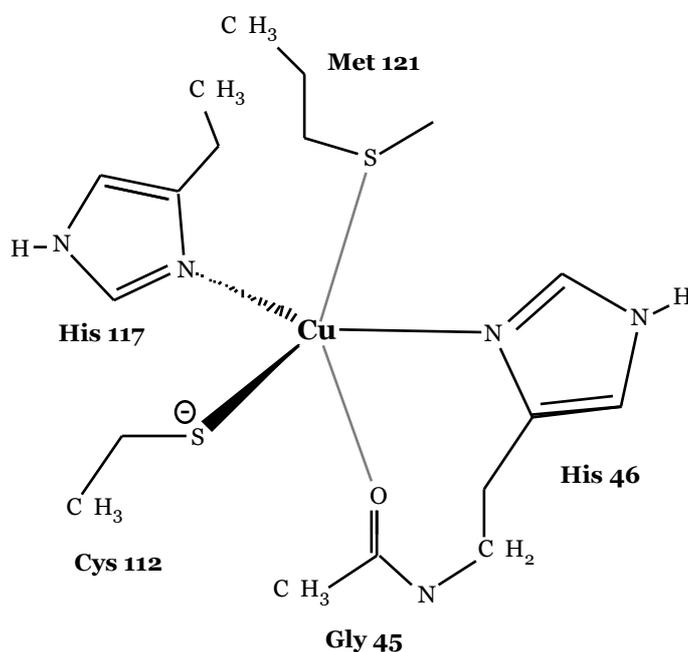


FIGURE 1.2.1. ACTIVE SITE OF AZURIN

The apoprotein structure⁷⁹ 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 Asn47⁸⁴, 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 mutant⁸⁴, 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 position³. 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) azurin⁴² 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 studies⁹⁶ 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 protein²⁴. 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 ligands¹⁰⁷. 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 calculations¹⁷ 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.

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.