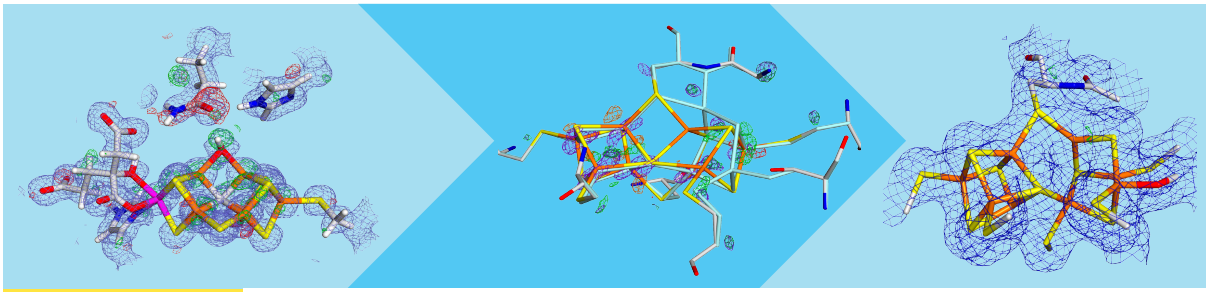
Computational Studies of Nitrogenase

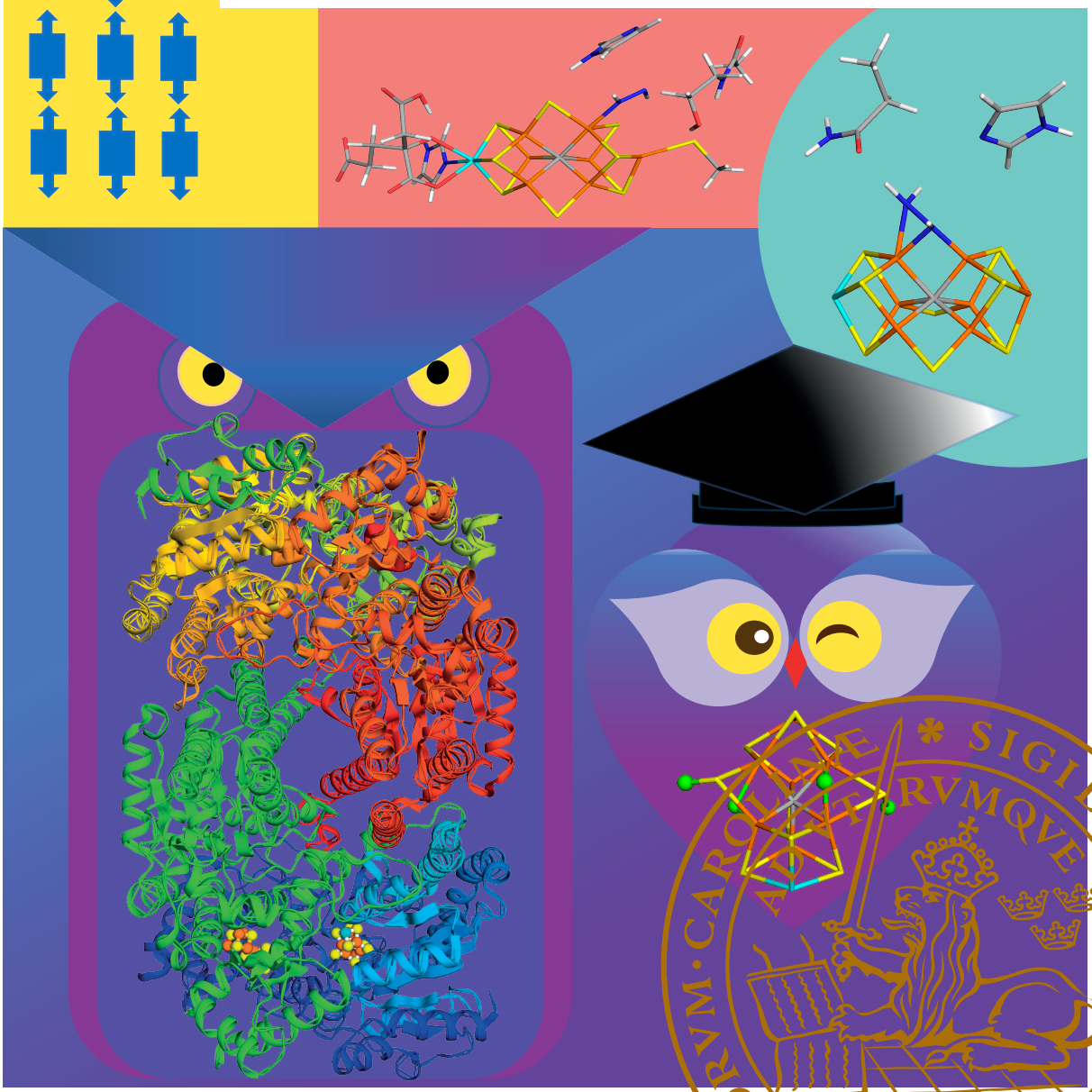
Cao, Lili

2020



Computational Studies of Nitrogenase

LILI CAO | DIVISION OF THEORETICAL CHEMISTRY | LUND UNIVERSITY



Computational Studies of Nitrogenase

Lili Cao

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Science, Lund University, Sweden.
To be defended at
Centre for Chemistry and Chemical Engineering, Hall A.
May 29th 2020, 13:00.

Faculty opponent
Dr. Ragnar Björnsson, Max-Planck-Institut für Chemische Energiekonversion,
Mülheim an der Ruhr, Germany

Organization	Document name
LUND UNIVERSITY, Centre for Chemistry and Chemical Engineering P.O. Box 124, 221 00, Lund, Sweden	Doctoral dissertation
	Date of issue 2020-05-29
Author Lili Cao	Sponsoring organization

Title and subtitle
Computational Studies of Nitrogenase

Abstract

Nitrogenase is the only enzyme that can convert the inert nitrogen molecule to ammonia, so that it can be used for biomass production and in biosynthetic pathways. It contains a complicated acitve site, composed of eight metal ions, nine sulfur ions and one carbide ion (the FeMo cluster). Although it has been thoroughly studied with crystallographic, biochemical, kinetic, spectroscopic and computational methods, the reaction mechanism is still not known and many conflicting hypotheses have been presented. To solve some of these problems, we have performed a thorough and systematic study of nitrogenase with various computational approaches, based on a combination of quantum mechanics (QM), molecular mechanics (MM) and sometimes also crystallographic refinment. We have:

- Decided the protonation states of eight key amino acid residues around the active site and showed that the homocitrate ligand is singly prontated on the hydroxide group.
- Studied how the the broken-symmetry (BS) state for the FeMo cluster depends on the QM method, the basis sets, the surrounding protein and the protonation and oxidation state of the cluster. Thereby, we could propse a practical prodecure to deal with these states in computational studies.
- Predicted the most stable protonation state for the E_0–E_4 states of nitrogenase with two density functional theory (DFT) methods: The TPSS functional perfers the protonation of Fe, whereas B3LYP prefers protonation of the carbide ion.
- Showed that different DFT methods give relative eneriges that can differ by ~1100 kJ/mol for nitrogenase. This is the main reason for the diverging computational results. Pure functionals and TPSSh predict the best geometries of the E_0 state, whereas B3LYP and PBE0 give more reliable H_2 dissociation energies.
- Showed that the most stable E_4 structure obtained with pure functionals has two hydride ions bridging between two pairs of iron ions and two protons on the sulfide ions, in agreement with experiments. We also found a new low-energy BS state in which only two Fe ions have minority spin.
- Predicted that the most stable binding mode of N_2H_2 to nitrogenase involves trans-HNNH binding to Fe2. However, other binding modes, e.g. involving cis-HNNH or NNH_2 and binding to Fe6, are rather low in energy.
- Suggested an alternating reaction mechanism for nitrogenase with a dissociated S2B ligand.
- Decided the most stable BS states for the P-cluster in four oxidation states and showed that the one-electron oxidised state involves a protonated Cys-88, but a deprotonated Ser-188.
- Developed a novel quantum-refinement approach allowing for disorder in the QM system and applied it to the P-cluster in two crystal structures of nitrogenase.
- Shown by quantum refinement that a recent crystal structure of V-nitrogenase does not involve a N-derived ligand, but rather a hydride-inhibited state with a significant amount of the undissociated S2B ligand.

Together, these studies have taken us a significant step closer to an atomic understanding of nitrogenase, showing that experiments and calculations start to converge, illustrating the strength of quantum refinement, but also pointing out important problems that need to be solved in computational studies of nitrogenase.

Key words nitrogenase, FeMo cluster, P-cluster, E_4, N fixation, QM/MM, quantum refinement, N_2 binding

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language** English
ISSN and key title		**ISBN** 978-91-7422-741-3
Recipient's notes	**Number of pages** 290	Price
	Security classification	

Signature  Date 2020-04-20

Computational Studies of Nitrogenase

Lili Cao

Front Cover by Lili Cao

Faculty of Science
Department of Theoretical Chemistry

ISBN 978-91-7422-742-0 (pdf)
ISBN 978-91-7422-741-3 (print)

Printed in Sweden by Media-Tryck, Lund University
Lund 2020

Table of Contents

List of Publications 7
List of Publications not Included in the Thesis 8
List of Article Contributions 9
Popular Science Summary 11

1. Introduction 17

2. Methods 19

2.1. Quantum Mechanical Methods 19
2.1.1. Hartree–Fock Theory 21
2.1.2. Basis Set 22
2.1.3. Density Functional Theory 24
2.2. Classical Methods 26
2.2.1. Molecular Mechanics 26
2.2.2. Molecular Dynamics Simulations 28
2.3. QM/MM 31
2.4. Crystallography 34
2.4.1. Crystallographic Refinement 38
2.4.2. Quantum Refinement 39

3. Nitrogenase 41

3.1. FeMo Cluster 42
3.1.1. Atomic Structure 42
3.1.2. Redox Properties 44
3.1.3. Reaction Mechanism 45
3.1.4. Previous Computational Studies 48
3.1.5. Broken-Symmetry States 49
3.2. P-cluster 51
3.2.1. Atomic Structure and Redox Properties 51
3.2.2. Electron-Transfer Reactions 53
3.2.3. Electronic Structure 54

4. Summary of the Articles 55
4.1. Paper I 56
4.2. Paper II 60
4.3. Paper III 63
4.4. Paper IV 67
4.5. Paper V 69
4.6. Paper VI 72
4.7. Paper VII 74
4.8. Paper VIII 76
4.9. Paper IX 78
4.10. Paper X 82
5. Conclusions and Outlook 85
References 91
Acknowledgments 99

List of Publications

Paper I: L. Cao, O. Caldararu & U. Ryde; "Protonation states of homocitrate and nearby residues in nitrogenase studied by computational methods and quantum refinement". J. Phys. Chem. B 2017, 121, 8242–8262.

Paper II: L. Cao & U. Ryde; "Influence of the protein and DFT method on the broken-symmetry and spin states in nitrogenase". Int. J. Quant. Chem. 2018, 118, e25627.

Paper III: L. Cao, O. Caldararu & U. Ryde; "Protonation and reduction of the FeMo cluster in nitrogenase studied by quantum mechanics/molecular mechanics (QM/MM) calculations". J. Chem. Theory Comput. 2018, 14, 6653–6678.

Paper IV: L. Cao & U. Ryde; "Extremely large differences in DFT energies for nitrogenase models". Phys. Chem. Chem. Phys. 2019, 21, 2480–2488.

Paper V: L. Cao & U. Ryde; "What is the structure of the E_4 intermediate in nitrogenase?" J. Chem. Theory Comput., 2020, 16, 1936–1952.

Paper VI: L. Cao & U. Ryde; "N_2H_2 binding to the nitrogenase FeMo cluster studied by QM/MM methods", J. Biol. Inorg. Chem., 2020, 25, DOI: 10.1007/s00775-020-01780-5.

Paper VII: L. Cao & Ulf Ryde; "Putative reaction mechanism of nitrogenase after dissociation of a sulfide ligand", J. Catalysis, submitted.

Paper VIII: L. Cao, M. Börner, J. Bergmann, O. Caldararu & U. Ryde; "Geometry and electronic structure of the P-cluster in nitrogenase studied by combined quantum mechanical and molecular mechanical calculations and quantum refinement". Inorg. Chem. 2019, 58, 9672–9690.

Paper IX: L. Cao & U. Ryde (2020) "Quantum refinement with disorder; applications to the P-cluster in nitrogenase", Acta Cryst. D, submitted.

Paper X: L. Cao, O. Caldararu & Ulf Ryde; "Does the crystal structure of vanadium nitrogenase contain a reaction intermediate? Evidence from quantum refinement", Acta Cryst. D, submitted.

List of Publications not Included in the Thesis

1. L. Cao, O. Caldararu, A. C. Rosenzweig & U. Ryde; "Quantum refinement does not support dinuclear copper sites in crystal structures of particulate methane monooxygenase", *Angew. Chem. Int. Ed.* **2018,** *57*, 162–166.
2. L. Cao & U. Ryde; "On the difference between additive and subtractive QM/MM calculations", *Front. Chem.* **2018,** *6*, 89.
3. G. Dong, L. Cao & U. Ryde; (2018) "Insight into the reaction mechanism of lipoyl synthase: A QM/MM study", *J. Biol. Inorg. Chem.* **2018,** *23*, 221–229.
4. V. Ekberg, L. Cao, M. M. Misini Ignjatović, M. A. Olsson & U. Ryde; "Attempts to improve alchemical relative binding free-energy simulations by quantum mechanical charges", *manuscript*.

List of Article Contributions

Paper I: I started the project by setting up the protein. I performed all the MD simulations, QM-cluster calculations, QM/MM, QM/MM-PBSA, QM/MM-2QM and QTCP calculations. I also performed the quantum refinement calculations and participated in the analysis. I participated in writing the manuscript.

Paper II: I performed all the calculations in this paper. I participated in writing the manuscript.

Paper III: I performed all the QM/MM calculations. I performed the quantum-refinement calculations and participated in the analysis of them. I participated in writing the manuscript.

Paper IV: I performed all calculations in this paper. I participated in writing the manuscript.

Paper V: I took part in design the project. I performed all calculations in this paper. I wrote the first draft of the manuscript.

Paper VI: I took part in design the project. I performed all calculations in this paper. I wrote the first draft of the manuscript.

Paper VII: I took part in design the project. I performed all calculations in this paper. I participated in writing the manuscript.

Paper VIII: I supervised a master student on the project, performed most of the QM/MM calculations in this paper and half of the quantum-refinement calculations. I participated in writing the manuscript.

Paper IX: I performed all calculations in this paper. I participated in writing the manuscript.

Paper X: I performed most of the calculations in this paper. I participated in writing the manuscript.

Popular Science Summary

Chemical elements are closely related to our daily life. In fact, our bodies contain many elements from the periodic table which are essential to ensure the proper function of life. As 65%–90% of each cell in the body is made up of water, oxygen and hydrogen are among the main components of the human body. The most abundant element in the human body is oxygen, but what is the most abundant element in the air? It is definitely not oxygen, otherwise a match would burn out swiftly. By volume, nitrogen constitutes 78% of the Earth's atmosphere, whereas oxygen is 21%. Interestingly, when nitrogen was first translated into Chinese, it was called "light air", which means that it "diluted" the oxygen in the air. The French chemist Antoine Laurent Lavoisier named nitrogen "azote", meaning "no life", because inhaling nitrogen causes suffocation. But with the deepening of research, scientists have discovered that nitrogen is not only indispensable for the human body, but also an essential element in the biosphere. It is clearly a most important gas.

🤔: *What happens if there is no nitrogen?*

👼: *Then there is no human…*

Nitrogen is one of the obligatory elements of amino acids, which serves as building blocks for the proteins. The amino acids arrange themselves in three dimensions to form proteins with special functions in cells, which in turn are used to build the various organs of the human body. The structures of proteins are very complicated, which endow them with fascinating functions, allowing them to perform a very large number of important physiological functions in the body. In fact, all life activities are carried out in an orderly manner thanks to the regulation by various functional proteins.

First, proteins form a major part of the body tissues. The growth and development of the body, the renewal and the repair of aged and damaged tissue all require the use of protein as one of the most important building materials.

Moreover, proteins constitute many physiologically active biomolecules, such as enzymes, hormones and antibodies. All physiological metabolism and chemical reactions in the human body are controlled by enzymes, which behave like chemical factories. The physiological functions are regulated by hormones, which often are proteins, such as growth hormone and insulin. The antibodies are an "assault team" that is active in the blood and has the function of protecting the body from bacteria and viruses and improving the body's resistance.

In addition, proteins can regulate the osmotic pressure. The protein content in blood plasma ensure that water maintained in balance between plasma and interstitial fluid. Lack of protein in the diet for a long time may reduce the protein

content in the plasma and water in the blood will penetrate into the surrounding tissues, leading the occurrence of nutritional oedema.

Furthermore, proteins can be an energy supply. Although this is not the main function of protein, in the absence of other energy sources, proteins can also be used to generate energy.

Proteins are involved in all human activities. There is no life without proteins and there is no protein without nitrogen. All this shows how important nitrogen is to humans.

🤔: *Want to get more "nitrogen"?*

😊: *Eat more legumes!*

If you think that humans get nitrogen through breathing, you are wrong. Instead, people have to get it through food. We eat food, which end up in the stomach and after a series of digestion and absorption, nitrogen is taken up by the body and participates in human life activities. The same applies to all animals, so the nitrogen ultimately comes from the plants.

Plants grow by converting solar energy into bioenergy through photosynthesis. The most important actor in this transformation process is the chloroplast. In the chloroplasts, there are many enzymes involved in the reactions. These enzymes are the most critical substances to promote the smooth progress of photosynthesis. They are proteins. Their production requires nitrogen, which determines the progress of photosynthesis and finally the quality of plant growth.

🤔: *So where does the nitrogen in plants come from?*

😊: *From the soil.*

Just as humans cannot directly obtain nitrogen from the atmosphere through breathing, the same applies to the plants. The majority of atmosphere is composed of nitrogen, but this is in the form of molecular nitrogen, N_2, which is incredibly inert. The triple bond in N_2 is so strong that plants or animals cannot break it. The plants can only utilise reduced or oxidised forms of this element, such as ammonia, nitrite and nitrate, taken up by the roots underground. The amount of these bioavailable forms of nitrogen in the soil is limited. These components are water soluble and they are easily washed away by rainwater. At the same time, some microorganisms in the soil can decompose these nitrogen compounds into nitrogen gas. Therefore, atmospheric nitrogen must be converted into the reduced or oxidised compounds to maintain the effective nitrogen content in the soil. The conversion of atmospheric N_2 into the nitrogenous compounds is called the nitrogen fixation.

 Lightning may cause a fire in a dry summer, but you may not know that lightning can also help plant growth. The heat and the energy within a lightning bolt induce chemical reactions that makes nitrogen readily available, acting as nature's natural fertiliser. But overall, the help from lightning is limited.

 Fritz Haber, a German chemist, developed an artificial nitrogen fixation procedure that converts atmospheric nitrogen into ammonia in an industrial process that requires high temperature and pressure. The produced ammonium can be applied to the soil and be absorbed by plants. This procedure is called the Haber–Bosch process. For this invention, Haber was awarded the Nobel Prize in Chemistry in 1918. This process has been essential to feed Earth's growing population and it is responsible for the exponential population boom. It replaced the passive situation of relying on natural nitrogen fertiliser and accelerated the development of world agriculture. Today, it is responsible for the food production for around half of the world's population. It is one of main industrial procedures and it stands for 1–2% of the total human energy consumption. In 2018, the global consumption of agricultural fertiliser was above 200 million tons, of which more than half is nitrogen.

Humanity's strong reliance on the Haber–Bosch process has generated some serious complications for the environment. For example, only a part of the fertilisers is taken up by the crop, whereas the remainder flows into waterbodies, upsetting nature's balance. The enrichment of nitrogen and other nutrients in the water causes rapid reproduction of algae and other plankton, which reduces the dissolved oxygen content in bottom waters, causing massive mortality of fish and other aquatic organisms. This phenomenon is referred to eutrophication. Furthermore, the leakage of nutrients to adjacent areas leads to overfertilisation of the landscape, changing the wildlife and threating many plants and animals by extinction. Moreover, the carbon dioxide emissions as a result of burning fossil fuels, in order to provide a large amount of energy to sustain the Haber–Bosch process, eventually lead to global warming. Therefore, we cannot continue to depend on this kind of nitrogen fertilisers and it is needed to find efficient alternatives for the Haber–Bosch process to sustain Earth's ever-growing population.

 Apart from the bioavailable nitrogen supplied by lightning and the Haber–Bosch process, there is a third way of fixing nitrogen. A special group of bacteria has the capacity to convert atmospheric nitrogen into ammonia. This is done by a special enzyme, called nitrogenase, which is the main player of this thesis.

The evaluations of Fritz Haber have been varying.

During World War I, Haber developed chlorine gas, mustard gas and other poisonous gases for warfare, causing nearly one million casualties.

As a consequence, people have either praised him as an angel who brings harvest or as a devil who brings disaster, suffering and death.

He was married to Clara Immerwahr, who in 1900 was the first woman in Germany to gain a doctorate. She committed suicide after her husband's first successful attack with chlorine gas.

This process is most well-known for the legumes, associating with a bacterium from the soil, called *Rhizobium*. On the roots of the legumes, special organs are found, called nodules. In these nodules, the *Rhizobium* bacteria are found and there they convert nitrogen into ammonia. This process is very energy intensive. The legumes supply the energy in the form of carbohydrates, which they have synthesised from the sunlight through photosynthesis. In return, the bacteria provide the legume a steady source of fixed nitrogen, which can be used to make proteins and chlorophyll. Thus, the legumes and *Rhizobium* form a symbiotic relationship for mutual benefit. When the legumes die, their leaves and roots decompose, and the remaining nitrogen is supplied to the soil. The key advantage of this process is that it does not require the high temperature and high pressure as in Haber–Bosch procedure. Instead it takes place at room temperature and normal pressure, using sunlight to drive it.

The use of legumes is important in agriculture. It has long been known that plant growth requires nutrients like nitrogen, but there is only a limited amount of nitrogen in the soil and once the crop has used it up, it has to be replenished. To avoid over-consumption of certain soil nutrients by repeated plantings of the same crop, farmers adopt the method of crop rotation. To return nitrogen to the soil, they plant legumes such as soybeans, clover, alfalfa, peas or peanuts every second year. Rotating crops by planting legumes enable farmers to avoid chemical fertilisers and to maintain or improve the soil quality at the same time.

Owing to the importance of agriculture and the environmental implications of nitrogen fixation, nitrogenase has attracted great interest from scientists. Naturally, the ultimate goal is to mimic the action of nitrogenase in an industrial

process, performed at ambient pressure and temperature. To this end, it is necessary to understand how the enzyme works. Although the atomistic structure of the enzyme is known and it has been thoroughly studied by both experimental and computational methods during more than forty years, there is still no consensus how the enzyme works.

Chemistry can not only be studied in an experimental laboratory. The latest 30 years, computational chemistry has established itself as a competitive alternative to experiments for the study of chemical reactions. It allows the computer simulate virtual scenarios close to real experiments. Calculations performed in the virtual lab could help researchers to understand, predict and discover new chemical phenomena and physical essences, avoiding restrictions in traditional chemical laboratory equipment or reagents. The foundation of computational chemistry is based on physical theories such as classical mechanics and quantum mechanics, using computers to perform a large number of numerical operations to explore the properties of chemical systems.

Computational chemistry is becoming more and more used in many fields. Typical applications include molecular structure analysis, drug design, study of enzyme-catalyzed reaction mechanism, and so on.

The versatility of proteins has stimulated scientists' interest in understanding the function of proteins on an atomic level, to get a better understanding of the structure–function relationship and ultimately guide scientists to design proteins with special functions. Usually, the protein structure is determined through X-ray diffraction. However, the process from a diffraction pattern to the prediction of a molecular model is complex and involves computational methods. We have developed a useful method, called quantum refinement, which is a combination of crystallographic refinement and quantum-mechanical calculations, to extract more information from the data and to better understand what is really seen in the structure.

Enzymes are catalytic proteins produced by living cells. A fundamental role of proteins is to behave as biological catalysts that speeds up chemical reactions at the mild conditions in living cells. To catalyse a certain reaction, a substrate needs to bind to a certain place of the enzyme which contains the specific catalytic function. This binding site is called the active site. Due to the high efficiency of the catalysis, the reactive states formed during the catalytic process are very short-lived, which make them difficult to study with experimental methods. Fortunately, computational chemistry can overcome these shortcomings and provide detailed information of entire reaction mechanisms at an atomic level. As we already know, proteins are made up of amino acids and the active site gets its specific function from the certain arrangement of these amino acids in 3D space. Thus, the study of reaction mechanisms should focus not only on the active site, but also consider the influence of the surrounding protein and solvent.

There is no doubt that a clear understanding of the structure–function relationship and detailed reaction mechanism of enzymes can help people to understand more about the function of proteins, understand what is important for the specific function and its regulation. This would open for the possibility to intentionally modify the structure and catalytic role of enzymes, or to design and synthesise artificial enzymes.

In this thesis, we have tried to understand the mechanism of nitrogenase with computational methods. This is a formidable task, because nitrogenase is one of the most complicated enzymes known. Moreover, previous computational studies have given varying and often contradicting results. Therefore, we have performed the study in many small steps, solving all problems involved and trying to reduce the many possibilities in an accurate and systematic way. We have decided the energetically most favourable protonation state of the first four states of the reaction and the best binding mode of N_2H_2. We have suggested a possible reaction mechanism of the enzyme. Moreover, we have developed an improved method to refine crystal structures and used this to discuss what is really seen in some crystal structures of nitrogenase. Altogether these studies provide enhanced understanding of the reaction mechanism of nitrogenase.

1. Introduction

Metabolism is the basis for all life activities of organisms and the metabolism is inseparable from the catalysis of enzymes. Most enzymes are special proteins with catalytic functions, also called biocatalysts, and almost all biochemical reactions are catalysed by enzymes. Therefore, a fundamental understanding of the processes of life must be founded on a detailed knowledge of how enzymes work, to extend the application of enzymes in chemical, food, healthcare, cosmetic, agriculture and detergent industry, and to design mutants with favourable properties.

Traditionally, the knowledge of the protein atomic structure has been deduced from X-ray crystallography, by interpreting the scattered pattern of X-rays from the crystalline structure. However, these reflections do not contain all information needed to determine the atomic structure, since the phase cannot be measured by experiments and hydrogen atoms are normally not discerned. The interpretation of the crystallographic data is facilitated by employing a priori chemical knowledge, which is integrated into the interpretation in the form of empirical restraints, similar to a force field employed in the computational chemistry. But such information is usually missing for non-standard ligands and metal ions. Moreover, chemical reactions are usually so fast that it is difficult to obtain any detailed information about the reaction mechanism by experimental means alone.

Fortunately, the explosive development of computers has allowed theoretical methods to become a competitive complement to experiments. The atomic models determined by crystallographers can be improved by introducing quantum chemistry in the refinement. Moreover, based on more accurate atomic structures, computational chemists can predict a detailed view of entire reaction mechanisms at the atomic level, by determining the structure of the intermediates and transition states of the reaction, and finding the rate-limiting step of the reaction. Computational chemistry has the advantage of directly providing energies (which govern chemical processes) and it is not subject to experimental limitations, such that key intermediates are invisible, short-lived, expensive or hazardous. In addition, computational chemistry is cheap and environment-friendly.

At least one third of all enzymes contain a metal ion as an essential cofactor. Metal ions are used especially for hard reactions involving small and

inert substrates like N_2 and H_2; therefore, they have a great industrial impact. Consequently, they have also attracted much interest from the computational community, in particular since the introduction of density functional theory (DFT). Systems of several hundreds of atoms can now routinely be studied by DFT methods with an accuracy approaching that of experiments.

One of the most important and most complicated enzymes is nitrogenase. It is the only enzyme that can cleave the triple bond of gaseous N_2, forming two molecules of ammonia and making atmospheric nitrogen available for biological systems.[1–3] Nitrogenase has been extensively studied with both experimental and computational methods,[3–5] but no consensus has yet been reached for the reaction mechanisms. In particular, previous computational studies have given very diverging and contradictory results.[4]

The aim of this thesis is to try to solve these problems and to provide an atomistic understanding of the reaction of nitrogenase. All the investigations have been based on various computational methods. In particular, we employed quantum mechanics, combined with molecular mechanics and X-ray crystallography refinement. We employ a systematic approach, trying to solve all problems involved and to study as many alternatives as possible. Although the time has not been enough to provide a full detailed mechanism of the reaction, we have solved several problems and pinpointed other, previously overlooked problems. By taking small, but significant steps towards the final goal, we for the first time start to see a convergence between experimental and computational approaches for nitrogenase.

2. Methods

In this section, I shortly describe most of the methods employed in the thesis.

2.1. Quantum Mechanical Methods

Quantum mechanical (QM) methods describe molecular energetics by solving the time-independent Schrödinger equation, which can be written in the form

$$\hat{H}\,\Psi = E\,\Psi \tag{2.1}$$

Mathematically, this is a differential equation. Thus, Ψ is an eigenfunction of the differential operator $\hat{H}$ and E is the corresponding eigenvalue. Ψ is called the wave function and it is a function of the coordinates of all electrons and nuclei in the system of interest. The system is completely described by the wave function, which means that we from the wave function can calculate any measurable properties of the system. However, it is often not possible to know the instantaneous properties of the system, but only the average value and probabilities of different properties. According to Born's interpretation, the wave function itself is a probability amplitude with no direct physical meaning, whereas $|\Psi|^2$ represents a probability density. $\hat{H}$ is the Hamilton operator, which defines the system by describing what (kinetic and potential) energy terms apply. Finally, E is the total energy of the system.

The Hamiltonian for a molecular system is given by

$$\hat{H} = \hat{T}_{\mathrm{e}} + \hat{T}_{\mathrm{n}} + \hat{V}_{\mathrm{en}} + \hat{V}_{\mathrm{ee}} + \hat{V}_{\mathrm{nn}} \tag{2.2}$$

where

$$\hat{T}_{\mathrm{e}} = -\frac{\hbar}{2m_e}\sum_{i=1}^{N}\nabla_i^2 \qquad\qquad \hat{T}_{\mathrm{n}} = -\frac{\hbar}{2}\sum_{A=1}^{n}\frac{1}{M_A}\nabla_A^2$$

$$\hat{V}_{\mathrm{en}} = -\frac{e^2}{4\pi\varepsilon_0}\sum_{\mathrm{i}=1}^{N}\sum_{\mathrm{A}=1}^{n}\frac{Z_{\mathrm{A}}}{|r_i - R_A|} \qquad \hat{V}_{\mathrm{ee}} = \frac{e^2}{4\pi\varepsilon_0}\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{|r_i - r_j|}$$

$$\hat{V}_{\mathrm{nn}} = \frac{e^2}{4\pi\varepsilon_0}\sum_{A=1}^{n}\sum_{B>A}^{n}\frac{Z_A Z_B}{|R_A - R_B|}$$

$$\nabla^2 = \left\{\frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2} + \frac{\partial}{\partial z^2}\right\} \tag{2.3}$$

where $\hbar$ is Planck's constant, m_e is the mass of the electron, e is the unit charge, ε_0 is the permittivity of vacuum, i and j refer to the electrons, A and B refer to the nuclei with the atomic numbers Z_A and Z_B. N is the total number of electrons and r_i their positions, whereas n is the number of nuclei, R_A their positions and M_A their masses. The operator ∇^2 is called the Laplacian, which represents the sum of the second derivatives with respect to the three Cartesian coordinates. $\hat{T}_{\mathrm{e}}$ and $\hat{T}_{\mathrm{n}}$ are the kinetic energy operators for the electrons and the nuclei. $\hat{V}_{\mathrm{en}}$, $\hat{V}_{\mathrm{ee}}$ and $\hat{V}_{\mathrm{nn}}$ represent the three Coulomb terms: The operator $\hat{V}_{\mathrm{en}}$ describes the attractive potential energy between the electrons and nuclei, whereas the operators $\hat{V}_{\mathrm{ee}}$ and $\hat{V}_{\mathrm{nn}}$ describe the electron–electron and nuclear–nuclear repulsive potential energy, respectively.

In practice, the Schrödinger equation can only be solved analytically for very few systems, such as hydrogen or hydrogen-like atoms. Unfortunately, for systems that involve more than two particles that interact with one another, analytical solutions become impossible. Therefore, approximations and simplifications must be introduced. One fundamental approximation that all calculations in this thesis have employed is to separate the motions of the electrons and the nuclei. This is called the Born–Oppenheimer approximation. It is based on the large mass difference between electrons and nuclei (a factor of 10^3–10^5) which indicates that the electron motions are instantaneous, whereas nuclei are stationary from the electronic point of view. Within this approximation, we only need to solve Eqn. 2.1 with the electronic Hamiltonian operator which can be written as

$$\hat{H}_{\mathrm{e}} = \hat{T}_{\mathrm{e}} + \hat{V}_{\mathrm{en}} + \hat{V}_{\mathrm{ee}} + \hat{V}_{\mathrm{nn}} \tag{2.4}$$

i.e., the nuclear kinetic energy $\hat{T}_{\mathrm{n}}$ is omitted. $\hat{H}_{\mathrm{e}}$ depends only on the positions of nuclei (via $\hat{V}_{\mathrm{en}}$ and $\hat{V}_{\mathrm{nn}}$) but not on their motions, and the resulting electronic wave function depends parametrically on nuclei coordinates.

2.1.1. Hartree–Fock Theory

To solve the electronic Schrödinger equation efficiently for a many-electron system, we need to rely on additional approximations, like neglecting magnetic and relativistic effects. Moreover, the simplest approaches also assume that the total wave function, which is a function of the coordinates of all electrons, can be factorised into a product of one-electron wave functions:

$$\Psi(r_1, r_2, \dots, r_N,) = \Phi_1(r_1)\Phi_2(r_2) \dots \Phi_N(r_N) \tag{2.5}$$

This is called the orbital approximation and is quite crude as it essentially assumes that each electron moves independently of all the other electrons, which of course is not true as all electrons repel each other.

The electron is known to have a half-integer spin with a spin quantum number of ½. In the presence of a magnetic field, the electron can align either along or opposite to the field. The two possible states are denoted by two spin eigenfunctions, denoted as α and β, and they are orthonormal. According to the Pauli principle, the total electronic wave function must be antisymmetric, meaning that interchanging the coordinates of two electrons should change the sign of wave function. For a many-electron system, these requirements can be achieved by building the wave function from a Slater determinant (SD)

$$\Phi_{\mathrm{SD}} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \dots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \dots & \phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) & \dots & \phi_N(N) \end{vmatrix} \tag{2.6}$$

where ϕ_i are the one-electron wave functions, spin-orbitals, given as a product of a spatial orbital and a spin function (α or β). For a molecular system, the one-electron wave functions in the determinant are the one-electron molecular orbitals (MOs).

Next, it is assumed that the one-electron MOs can be expressed as a linear combination of a set of known mathematical functions, called atomic orbitals (χ_j), the MO–LCAO approximation:

$$\phi_i = \sum_{j=1}^{M} c_{ij}\chi_j \tag{2.7}$$

Then, the variational principle can be applied, which states that an approximate wave function always gives a too high energy, compared to the exact solution. This means that the c_{ij} expansion coefficients in Eqn. 2.7 can be determined by minimising the energy, an optimisation problem that is ideal for computers.

Minimising the energy of a Slater determinant leads to definition of the Fock operator, which is given by

$$\hat{F}_i = \hat{h}_i + \sum_j^N (\hat{J}_j - \hat{K}_j) \tag{2.8}$$

where

$$\hat{h}_i = -\frac{\hbar}{2m_e}\nabla_i^2 - \frac{e^2}{4\pi\varepsilon_0}\sum_{A=1}^N \frac{Z_A}{|r_i - R_A|} \tag{2.9}$$

describes the kinetic energy of electron i and the attraction to all the nuclei. The electron repulsion to all the other electrons is described by Coulomb ($\hat{J}_j$) and exchange ($\hat{K}_j$) operators. The MO orbital energy and the total energy can be written as

$$\varepsilon_i = \langle \phi_i | \hat{F}_i | \phi_i \rangle = \hat{h}_i + \sum_j^N (\hat{J}_{ij} - \hat{K}_{ij}) \tag{2.10}$$

$$E = \sum_i^N \varepsilon_i - \frac{1}{2}\sum_{ij}^N (J_{ij} - K_{ij}) + V_{nn} \tag{2.11}$$

It can be seen that the total energy, E, is not simply a sum of all the MO orbital energies, because the Fock operator also contains the interaction terms introduced by $\hat{J}$ and $\hat{K}$. The energy is also not the exact energy because of the absence of electron correlation in HF calculations, i.e. the instantaneous repulsion between the individual electrons. Instead, each electron is interacting with the average field of all the other electrons. Thus, the HF method is a mean-field approximation.

2.1.2. Basis Set

The atomic orbitals in the MO–LCAO approach (Eqn. 2.7) are often called the basis functions or the basis set. The MO–LCAO expansion is a convenient way to solve a differential equation and it is in principle no approximation if the number of basis functions is infinite, but this is of course impossible in real calculations. Since only a limited number of basis functions are used, the size and quality of the basis set are important for the accuracy of QM calculations.[6] The basis functions should be selected so that they are as similar as possible to the real atomic wave functions and so that the involved integrals can be calculated rapidly.

There are two types of basis functions: Gaussian type orbitals (GTO) and Slater type orbitals (STO). The GTOs have the form

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{2n-2-l}e^{-\zeta r^2} \quad (2.12)$$

whereas the STOs have the form

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{n-1}e^{-\zeta r} \quad (2.13)$$

In both equations, n, l, m are the quantum numbers. $Y_{l,m}(\theta,\varphi)$ are the spherical harmonics functions, r is the distance between the electron and the nuclei. ζ is a predetermined constant. A small ζ gives more diffuse functions, whereas a large ζ gives a tighter function. Diffuse functions are important to describe anions or excited states, which often have loosely bound electrons and they are also important to describe properties that depend on the tail of wave function. Tight functions are important for the core electrons.

The GTOs, in contrast to the STOs, do not have a correct shape of the wave function, compared to the analytic solutions for hydrogen-like atoms. The difference is especially pronounced close to the nucleus and at large distances, where they fall off too rapidly. Therefore, more GTOs than STOs are needed to achieve the same accuracy. However, this is more than compensated by the fact that integral calculations are much more effective with GTOs. Consequently, the great majority of QM software employs GTOs.

The smallest basis set employs just enough functions for a minimum description of the occupied orbitals, i.e. one basis function per electron pair (assuming a closed-shell system), which is called a single zeta (SZ) basis set (zeta refers to the ζ term in Eqn. 2.13). An improvement of SZ is a double zeta (DZ) basis set, which employs two basis functions for each electron pair. Likewise, triple zeta (TZ) basis sets contain three basis functions for each electron pair. In practice, it is usually enough to have only a single basis function for the core orbitals, but two or three for the valence orbitals (because only those orbitals change significantly upon formation of chemical bonds). This is marked by adding the word "valence" before the basis set, e.g. VDZ (valence double zeta, also called split valence basis sets, SV) and VTZ.

Polarisation functions are basis functions with a higher angular-momentum quantum number than that of the valence shell in the atom they describe. They are added to treat charge polarisation effects and they are needed to get reasonable energies and geometries, especially when electron correlation is considered.

When optimising basis functions in terms of the energy, it turns out that it is important to get a good description of the core electrons, although they are less important for chemistry, which mainly depends on the valence electrons. However, core orbitals usually change very little in different molecules. Therefore, the core electrons can be described as a linear combination of several basis functions with fixed coefficients and the resulting basis functions are called

contracted GTOs (CGTOs). The number of the primary GTO involved in the linear combination is a compromise between the accuracy required and the gain in computation efficiency.

2.1.3. Density Functional Theory

Density Functional Theory (DFT) is based on the Kohn–Hohenberg theorem, stating that the electron density (ρ) uniquely determines the wave function and therefore the ground-state electronic energy.[7,8] This is highly remarkable, because the wave function is a function of $3n$ variables, where n is the number of particles (electrons and nuclei), whereas the electron density is a function of only the three Cartesian coordinates. Early efforts treated the system as non-interacting uniform electron gas, which is known as Thomas–Fermi theory. However, the electron kinetic energy was represented poorly, making this version of DFT unsuitable for chemistry.

The foundation for the use of DFT in computational chemistry is the orbitals introduced by Kohn and Sham.[9] The Kohn–Sham theory splits the electron kinetic energy into two parts, one of which can be solved exactly and the other is a correction term. The first part T_S, is calculated from Schrödinger equation, given as a Slater determinant, under the assumption that there are no electron interactions within the system. The correction part, which is the difference between the exact kinetic energy for a real system and T_S. This correction is small and is absorbed into the exchange–correlation term, E_{XC}. The general expression for DFT energy is

$$E_{DFT}[\rho] = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{XC}[\rho] \tag{2.14}$$

where T_S is the non-interacting electron kinetic energy, E_{ne} the attractive potential between nuclei and elections and $J[\rho]$ the Coulomb repulsion between electrons. The last term, E_{XC}, contains the kinetic correction, as well as the exchange and correlation energy. Unfortunately, the form of the E_{XC} is not known exactly. Therefore, extensive development of methods is directed to find a proper approximate form for this term.

The simplest model, local density approximation (LDA), treats the electron density as a uniform electron gas, i.e. a density that varies slowly. The exchange energy is then given by Dirac's formula

$$E_X^{LDA}[\rho] = -C_X \int \rho^{4/3}(r)dr \tag{2.15}$$

For an open-shell system, the exchange energy can be described by the local spin density approximation (LSDA),

$$E_{\mathrm{X}}^{\mathrm{LSDA}}[\rho] = -C_{\mathrm{X}} \int (\rho_{\alpha}^{4/3} + \rho_{\beta}^{4/3})\, dr \qquad (2.16)$$

The correlation energy of the uniform electron gas has been estimated by Monte Caro methods, modified by Vosko, Wilk and Nusair (VWN)[10] to apply in DFT calculations. LSDA usually gives quite poor results.

A significant improvement over LSDA is obtained by considering the system as a non-uniform electron gas and allowing the exchange and correlation energy to depend not only on the local electron density, but also on the gradient of the electron density. This approach is called the generalised gradient approximation (GGA). GGA strongly improved the DFT results and made it popular for chemical systems. Several GGA functionals were employed in this thesis, viz. BP86, BLYP, PBE, B97D and M06L.[11–16]

Meta-GGA methods additionally includes a dependence of the Laplacian of the electron density (i.e. the second derivative). They typically give results with an accuracy that is comparable to those obtained with GGA functionals. One meta-GGA method, TPSS,[17] was the main DFT functional used in this thesis, because it gives accurate geometries and reasonable energies at a modest computational cost.

A common approach to improve the DFT functionals is to include a fraction of non-local exact HF exchange into the exchange–correlation term. Functionals including such a term are called hybrid functionals, whereas those without it are often called pure functionals. Hybrid functionals can be expressed as an appropriate combination of the exact exchange, LSDA and GGA energies. For instance, the B3LYP functional

$$E_{\mathrm{XC}}^{\mathrm{B3LYP}} = (1-a)E_{\mathrm{X}}^{\mathrm{LDA}} + aE_{\mathrm{X}}^{\mathrm{HF}} + b\Delta E_{\mathrm{X}}^{\mathrm{B}} + (1-c)E_{\mathrm{C}}^{\mathrm{LDA}} + cE_{\mathrm{C}}^{\mathrm{GGA}} \qquad (2.17)$$

with $a = 0.20, b = 0.72, c = 0.81$, consists of 80% of LDA exchange, 20% of HF exchange, 72% of Becke 1988 gradient-corrected exchange,[11] 19% of LDA correlation and 81% of GGA correlation energy.[13] Hybrid functionals employed in this thesis are TPSSh, B3LYP, PBE0, M06, BHLYP, M06-2X and M06-HF, which contain 10, 20, 25, 27, 50, 54, and 100% HF exchange, respectively.[11,13,18–23]

The advantage of DFT methods is that the correlations problem is avoided. With similar size of basis set, DFT methods typically give much better results than HF at a similar cost in terms of computer time. DFT is also appropriate for large systems, since the time consumption is proportional to N^3, where N is the number of basis functions, whereas HF scales as N^4 and correlated methods have even worse scaling. In general, DFT methods give accurate geometries with errors of up to 0.05 Å for metal–ligand bond lengths.[24]

2.2. Classical Methods

In contrast to QM methods, electrons are not explicitly considered in classical methods. Instead, atoms are modelled as balls and the bonds between atoms are represented as springs. The energy of the system is described by a potential-energy function, which is a mathematical function of the atomic coordinates that gives the energy of the system. Thereby, the time-consuming solution of the Schrödinger equation is avoided. The parameters of this function are normally called a force field and methods employing such a classical energy function are called molecular-mechanics (MM) methods.

2.2.1. Molecular Mechanics

A typical potential-energy function for a biomolecular system involves five energy terms

$$U_{\text{total}} = U_{\text{bonds}} + U_{\text{angles}} + U_{\text{dihedrals}} + U_{\text{el}} + U_{\text{vdW}} \tag{2.18}$$

The first three terms describe covalent contributions, coming from all bonds, bond angles and dihedral angles. The last two terms describe the intermolecular interactions, coming from the electrostatic and Van der Waals interactions.

Typically, the bond stretching term is given by the first term of a Taylor series,

$$U_{\text{bonds}} = \sum_{\text{bonds } i} C_i (r_i - r_{i0})^2 \tag{2.19}$$

where C_i is the force constant, r_i is the actual bond length and r_{i0} is the corresponding ideal bond length at equilibrium. This formula is also known as Hooke's law. Modelling a bond as a harmonic oscillator might lead to inaccuracies, since the real bond stretching energy is not harmonic. Therefore, a more expensive Morse potential are used in some applications.[25]

The bond angle term is also often a harmonic potential,

$$U_{\text{angles}} = \sum_{\text{angles } i} D_i (\alpha_i - \alpha_{i0})^2 \tag{2.20}$$

where D_i is the force constant, α_i is the actual angle and α_{i0} is the corresponding ideal bond angle at equilibrium. It is notable that the angles subtended at an atom are not independent, which may lead to problems during the parameterisation.

The dihedral angle, defined as the angle between the two planes $a_1 - a_2 - a_3$ and $a_2 - a_3 - a_4$ defined by four covalently connected atoms $a_1 - a_2 - a_3 -$

a_4, describe the rotation around a bond. The potential energy is often expressed as

$$U_{\text{dihedrals}} = \sum_{\text{diheddrals } i} \sum_{j=1} E_{ij}\left(1 + \cos\left(j\varphi_i + \delta_j\right)\right) \tag{2.21}$$

where E_{ij} is the force constant, φ_i is the actual dihedral angle and δ_j is a phase shift. In variance to the bond and angles, the rotation around a bond is periodic, i.e. a rotation of 360° brings the bond back to the starting point again; therefore, a trigonometric function is used. *j* is the periodicity of the function and typically periodicities up to 4 or 6 are allowed.

To ensure planarity of conjugated groups and the correct stereochemistry of chiral centres, improper torsions are considered in many force fields. They can also be described by the angle between two planes $a_1 - a_2 - a_3$ and $a_3 - a_2 - a_4$, but for four atoms that are not linearly connected (instead three of the atoms are typically covalently bound to the fourth). They can be treated in the same way as a dihedral angle, but some force fields instead employ a harmonic function.

The fourth term is the Coulomb electrostatic interaction energy, expressed as

$$U_{\text{el}} = \sum_{\text{pairs } ij} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}} \tag{2.22}$$

where ε_0 is the permittivity of vacuum, ε is the dielectric constant of the medium (normally set to unity), q_i and q_j are the partial charges of the atoms and r_{ij} is the distance between them. The interaction between charged atoms is long-ranged, large and diminishes with the inverse distance (r^{-1}). A Coulombic energy term is quite crude, because it ignores electrostatic polarisation and models the charge distribution of the molecule by atomic point-charges, rather than the full electron density.

The last term describes the Van der Waals interactions, for which a Lennard-Jones potential is usually used:

$$U_{\text{vdW}} = \sum_{\text{pairs } ij} 4\varepsilon_{ij}\left[\frac{\sigma_{ij}}{r_{ij}^{12}} - \frac{\sigma_{ij}}{r_{ij}^{6}}\right] \tag{2.23}$$

Here, ε_{ij} is the minimum potential energy, σ_{ij} is the distance between two atoms at which the potential energy is zero and r_{ij} is the distance between two atoms. This interaction consists of two parts, dispersion and exchange repulsion. Dispersion arises from the instantaneous correlation between the electronic motions in two non-bonded atoms. It is always attractive. The exchange-repulsion reflects the nuclear–nuclear repulsion, the electron–electron repulsion and the Pauli principle. It is strongly repulsive and dominates all the other

interactions at very short distances, thereby avoiding that the atoms clash into each other. Both the electrostatic and Van der Waals interactions are normally ignored for atoms that are connected by a covalent bond or are two bonds apart, whereas they are scaled down for atoms that are three bonds apart (because such interactions are described by the three bonded terms).

All the parameters in the formulas above need to be determined before the method can be used. They can be obtained from experiments or from QM calculations. In this thesis, we employed the Amber ff14SB force field[26] for the protein, the general Amber force field (GAFF)[27] for small organic molecules and the TIP3P force field (transferable intermolecular potential with three interaction points) for water molecules[28]. Partial charges for new molecules and metal sites were obtained with the restrained electrostatic potential (RESP) method,[29] based on QM calculations.

2.2.2. Molecular Dynamics Simulations

Sometimes, it is of interest to know the dynamics of a molecular system, e.g. to sample possible configurations in order to estimate thermodynamic averages and free energies. This can be done by molecular dynamics (MD) simulations.

In MD, the atomic motions are assumed to obey Newton's equation of motion (i.e. classical mechanics). Since MD requires a large number of energy evaluations, the potential energy and the forces between the particles are usually calculated from a MM force field. The force is the negative of the gradient, i.e. the first derivation of the potential energy respect to the positions. Thus, for a given potential energy U, the force is

$$F(x(t)) = -\frac{dU(x(t))}{dx(t)} \tag{2.24}$$

where $x(t)$ is the position at time t. According to Newton's second law of motion, we have

$$F = ma(t) = m\frac{dv(t)}{dt} = m\frac{d^2x(t)}{dt} \tag{2.25}$$

where m is the particle, $v(t)$ is the velocity and $a(t)$ is the acceleration at time t.

The coordinate of a particle at time t can be expressed as a Taylor expansion,

$$x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{F(t)}{2m}\Delta t^2 + \ldots \tag{2.26}$$

therefore,

$$x(t + \Delta t) \approx x(t) + v(t)\Delta t \tag{2.27}$$

Similarly, we could also get

$$v(t + \Delta t) \approx v(t) + a(t)\Delta t \tag{2.28}$$

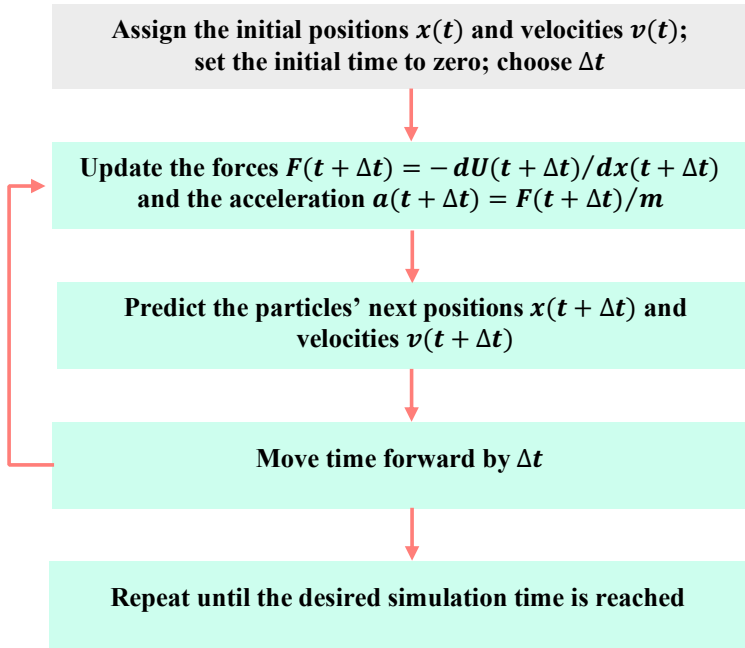
Thus, based on $x(t)$, $v(t)$ and $F(t)$ at time t, we can use Eqns. 2.25–28 to predict and move the particles to new positions $x(t + \Delta t)$ and update the velocities to $v(t + \Delta t)$ at time $t + \Delta t$. Since the potential energy of the system depends on the coordinates, a new set of forces and accelerations can then be derived by Eqns. 2.24 and 2.25. This can be repeated a large number of times, giving the time evolution of the molecular system, called trajectories. The same method is in principle used when calculating the trajectories of missiles or the movement of celestial objects.

In a real experiment measuring some properties, the properties are measured by some instruments during a certain time. To avoid statistical errors, the properties are averaged values over the time of measurement. The longer the average, the more accurate will the results be. A similar approach is applied with MD simulations. First, we assign initial velocities $v(t)$ and positions $x(t)$ to all particles in the studied system. Then we solve Newton's equation of motion until the properties of the system no longer change with time (i.e. until they have reached an equilibration). A simplified procedure is depicted in Scheme 2.1. Many other algorithms for MD than those in Eqn. 2.27–28 have been developed with slightly improved properties.[30]

The time step Δt needs to be smaller than the fastest atomic movement in the system and it is normally ~0.5 fs for a molecular system. The fastest movements are the bond vibrations involving hydrogen atoms. Since these are not normally interesting, it is common to constrain all bonds involving hydrogen atoms to their equilibrium values using the SHAKE algorithm.[31] Then, the time step can be increased to 2 fs. This does normally not affect the dynamics of the system, in contrast to constraints of angles and dihedrals.

In this thesis, MD simulations have been used to equilibrate atoms added to the crystal structure of nitrogenase (hydrogen atoms and solvation water molecules). MD has also been used to evaluate which protonation state of various protein residues reproduce the crystal structure best and to calculate protonation free energies in Paper I.

Scheme 2.1. A simplified MD scheme.



2.3. QM/MM

The combined QM and MM (QM/MM) method is an approach that is intended to combine the accuracy of QM calculations and the speed of MM calculations. For biomacromolcules like proteins, a small but interesting region (e.g. the active site) is modelled by QM calculations, whereas the rest of the protein is modelled by MM calculations.[32][33,34] In a typical QM/MM application, the system is usually divided into three subsystems, as shown in Figure 2.1. System 1 contains the atoms of prime interest (10–300 atoms) and it is represented by a wave function. System 2 consists of all atoms (from both the macromolecule and water) within 6–16 Å of any atom in system 1 (1000–10000 atoms) and is optimised by MM methods in each iteration of the geometry optimisation of system 1. System 3 includes the rest of the simulated system, typically including a significant amount of solvating water molecules and is also treated by MM methods. All atoms in this system are kept fixed in their starting coordinates, which typically are taken from a crystal structure.

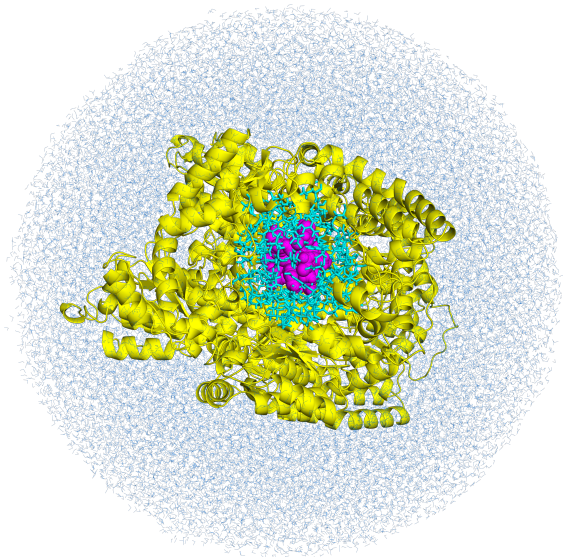


Figure 2.1: The systems in a QM/MM calculation: system 1 (shown as space-filling spheres in magenta), system 2 (sticks in cyan) and system 3 (rest of the protein in yellow and water in blue).

A simple energy function for a QM/MM approach is given by

$$E_{\mathrm{QM/MM}} = E_{\mathrm{QM1}} + E_{\mathrm{MM123}} - E_{\mathrm{MM1}} \tag{2.29}$$

where E_{QM1} is the energy of the system 1, calculated by a QM method, E_{MM123} is the energy of the total system calculated by MM and the last term, E_{MM1}, is the MM energy of system 1. E_{MM1} is needed to avoid double-counting of the energy of system 1. In this expression, all interactions between systems 1 and 2 are treated by the MM calculations. Therefore, no polarisation is considered, except within system 1. This approach is called mechanical embedding.

In our work, we used a more sophisticated approach, called electrostatic embedding. In this, all electrostatic interactions between the QM and MM regions are treated in the QM calculations. The energy formula is then expressed as

$$E_{\mathrm{QM/MM}} = E_{\mathrm{QM1+ptch23}} + E_{\mathrm{MM123},q_1=0} - E_{\mathrm{MM1},q_1=0} \qquad (2.30)$$

Here, system 1 is still represented by a wave function in the QM calculation, $E_{\mathrm{QM1+ptch23}}$, but all atoms in systems 2 and 3 are represented by an array of partial point charges, one for each atom, taken from MM setup. Thereby, the polarisation of the QM system by the surroundings is included in a self-consistent manner. In the MM calculations ($E_{\mathrm{MM123},q_1=0}$ and $E_{\mathrm{MM1},q_1=0}$), the electrostatic interactions involving system 1 are then turned off by zeroing the charge of QM region (to avoid double counting). This is normally considered as a more accurate approach, but the lack of polarisation of the MM system might lead to some overpolarisation of the QM system.

When there is a bond between systems 1 and 2 (a junction), a special treatment is needed, because QM calculations do not allow unfilled valences.[33] We have used the hydrogen link-atom approach: The QM system was capped with hydrogen atoms (hydrogen link atoms, HL), the positions of which are linearly related to the corresponding carbon atoms (carbon link atoms, CL) in the full system.[35,36] All atoms were included in the point-charge model, except the CL atoms.[37] Therefore, the employed energy function is:

$$E_{\mathrm{QM/MM}} = E^{\mathrm{HL}}_{\mathrm{QM1+ptch23}} + E^{\mathrm{CL}}_{\mathrm{MM123},q_1=0} - E^{\mathrm{HL}}_{\mathrm{MM1},q_1=0} \qquad (2.31)$$

where $E^{\mathrm{HL}}_{\mathrm{QM1+ptch23}}$ is the QM energy of the QM system truncated by HL atoms and embedded in the set of point charges modelling systems 2 and 3. $E^{\mathrm{HL}}_{\mathrm{MM1},q1=0}$ is the MM energy of the QM system, still truncated by HL atoms, but without any electrostatic interactions. Finally, $E^{\mathrm{CL}}_{\mathrm{MM123},q1=0}$ is the classical energy of all atoms in the system with CL atoms and with the charges of the QM region set to zero. There are other methods to treat junctions, e.g. the local self-consistent field [35] and the generalised hybrid orbitals approaches.[38]

QM/MM has been our main method, used in all my publications I–X. The reason for this is that it includes explicitly the entire protein in the calculations. Thereby,

the risk of biasing the results by the choice of the QM system is strongly reduced. The alternative is to use QM-cluster calculations, in which only a rather small QM system (100–300 atoms) is considered, whereas the remaining protein and solvent is modelled by a continuum-solvation model.[39] Restrictions in the geometry caused by the surrounding protein is normally modelled by keeping some atoms fixed, where bonds have been broken (but there is no model of the bulk of the protein surrounding the cluster). Test calculations have shown that QM/MM calculations converge faster than QM-cluster calculations with respect to the size of the QM system.[37,40–44] QM/MM also gives more reliable geometries.[41]

2.4. Crystallography

Atomistic models of proteins are essential tools for scientists to understand the function of proteins and to reveal the molecular details of life. The atomic structures of a protein can be determined by several methods, e.g. nuclear magnetic resonance (NMR), X-ray crystallography, neutron crystallography and cryo-electron microscopy. Almost all the published protein structures are deposited in the Protein Data Bank (PDB)[45] (https://www.rcsb.org) and to date, over 89% of the released structures are from X-ray crystallography experiments.

Protein crystallographic structure determination is a complicated process that involves several steps: growing crystals, recording diffraction patterns, processing data, obtaining the phases, model building, refinement and model validation.[46]

Growing crystals: Under certain circumstances, many proteins can form crystals. The resulting crystal is composed of highly ordered molecules, which can be imagined as an effective translational repetition of many identical unit cells. Within the unit cell, the asymmetric unit is the smallest volume element that contains the complete structural information of the protein. The growth of crystals requires highly purified protein and it is often a bottleneck in X-ray crystallography to find conditions under which the protein produces crystals of a sufficient quality.

Recording diffraction patterns: Once a crystal is obtained, it is exposed to an X-ray source and an X-ray detector is used to collect the scattered data. Nowadays, this is often done at a synchrotron, which produces powerful of X-rays. The X-rays interact with the crystal in a particular way, determined by the positions of the atoms and this phenomenon is known as diffraction. The diffracted X-rays form spots that are recorded by the detector and their intensities contain structural information about the protein. For each diffraction spot the diffracted rays satisfy Bragg's law,

$$2d_{hkl}sin\theta = n\lambda \tag{2.32}$$

where d_{hkl} is the interplanar spacing (h, k, and l are the Miller indices, specifying a certain reflection), θ is the angle between the plane and the incident X-ray, λ is the wavelength of the X-ray and n is an integer, representing the order of reflection. The intensities of the reflections on the detector depend on the distribution of electron density in the unit cell. After the X-ray data collection, the spots are associated with Miller indices in a process called indexing.

Processing data: The reflection with indices hkl can be described as a structure factor, F_{hkl}.[46] It is a three-dimensional periodic function that involves two

parameters, amplitude $|F_{hkl}|$ and phase (α_{hkl}). The amplitude is proportional to the square root of the intensity I_{hkl} of reflection hkl and can therefore be directly obtained from the measured reflection intensities. However, the phase of a F_{hkl} cannot be directly obtained through the X-ray scattering experiment.

On the other hand, the structure factor can be written as a sum of terms that describe the contribution from each atom in the unit cell to the reflection F_{hkl},

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j+ky_j+lz_j)} \tag{2.33}$$

where n is the number of atoms in the unit cell, x_j, y_j, and z_j are the Cartesian coordinates of atom j, whereas h, k and l are the Miller indices and $i = \sqrt{-1}$. f_j is the scattering factor of atom j. It depends on the element and the formal charge (number of electrons) and it is normally obtained by treating the atom as a sphere of electron density. This describes a method to obtain calculated amplitudes and phases if you have an atomic model of the protein.

Alternatively, F_{hkl} could also be obtained from the electron density in the unit cell

$$F_{hkl} = \int_V \rho(x,y,z)\, e^{2\pi i(hx_j+ky_j+lz_j)} dV \tag{2.34}$$

where V is the volume of the unit cell, $\rho(x,y,z)$ is the electron density at position (x,y,z). Owing to the reversibility of a Fourier transform, the electron density could also be written as a Fourier sum and depicted as a function of F_{hkl},

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l |F_{hkl}|\, e^{-2\pi i(hx+ky+lz)+i\alpha_{hkl}} \tag{2.35}$$

where $|F_{hkl}|$ is the amplitude and α_{hkl} is the phase. The aim of X-ray crystallography is to calculate $\rho(x,y,z)$ and build an atomic model of the protein into it. The prime problem is that for a new structure the phases are unknown.

Obtaining the phases: There are several ways to deal with this phase problem.[46] If there are already crystal structures of this protein (e.g. with different ligands) or of a similar protein, that structure can be used as an initial model of the protein in this structure (possibly after translation and rotation; called molecular replacement) and initial phases can be calculated from Eqn. 2.33 and used in Eqn. 2.35 to calculate the electron density. If there are no previous structures, other methods need to be used, e.g. isomorphous replacement with heavy-atom derivatives or anomalous scattering. Initial phases are typically incomplete and inaccurate, but enough to build an initial model, which then can be iteratively improved and used to calculate improved phases by Eqn. 2.33 Consequently, the phases and the electron density are biased by this initial model.

In crystallography, a resolution of 2 Å means that the analysis takes into account reflections out to 1/2 Å–1 from the centre of diffraction pattern in the reciprocal lattice. It refers to the amount of data used in the structure determination. A model refined to high resolution provides more detailed structural information than a low-resolution structure, as can be seen in Figure 2.2.

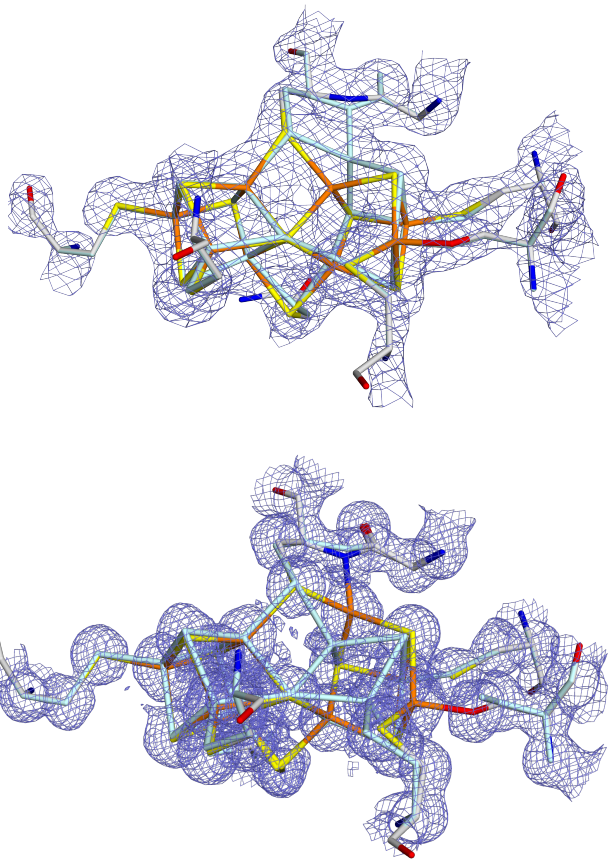


Figure 2.2: $2mF_o - DF_c$ electron-density maps of two nitrogenase crystal structures at resolutions of a) 2.1 Å, 6CDK [47] and b) 1.0 Å, 3U7Q[48]. Both maps are contoured at the 3 σ level.

Model building and refinement: Once the phases are determined, crystallographic software can calculate an electron density map from the experimental data, into which an atomic model can be built. Since the phases are rough estimates from molecular replacement, heavy-atom derivatives or anomalous scattering, the first map may be relatively uninformative. The quality of the map is then improved by an iterative process of crystallographic refinement

and model building.[46] The aim of the refinement is to produce a molecular model that optimises the agreement with the original reflection intensities.

In practice, the atomic model contains more parameters than the atomic coordinates. In particular, a typical model includes also atomic displacement parameters (ADPs; also called *B*-factors) and occupancies, which reflect the thermal and static disorder of the structure. With these, Eqn. 2.33 becomes

$$F_c = G \cdot \sum_j \; n_j f_j e^{2\pi i(hx_j+ky_j+lz_j)} \cdot e^{-B_j[(sin\theta)/\lambda]^2} \qquad (2.36)$$

where B_j is the atomic displacement parameter of each atom, describing the fluctuation of the atoms around their equilibrium positions specified in the model. n_j is the occupancy of each atom, another parameter included in the refinement. For most atoms, it is unity, indicating that it has only a single conformation. However, some groups may exhibit disorder and therefore several conformations, which each has $n_j < 1$.

Many graphical software are available to support the model building, e.g. Coot and Pymol. Several maps are used to support the model building, but the two most common are the $F_o - F_c$ and $2F_o - F_c$ maps. They are calculated as follows

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l (|F_o| - |F_c|)\, e^{-2\pi i(hx+ky+lz)+i\alpha_{hkl}} \qquad (2.37)$$

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l (2\,|F_o| - |F_c|)\, e^{-2\pi i(hx+ky+lz)+i\alpha_{hkl}} \qquad (2.38)$$

where $|F_o|$and $|F_c|$are the observed (experimentally measured) and the calculated structure factor amplitudes based on the current model, respectively. The $2F_o - F_c$ map is similar to the net electron density map in Eqn. 2.35, but the F_c term somewhat reduces the bias from the phases obtained from the current model. In a proper structure, all atoms in the model should be within the density of the $2F_o - F_c$ map (within the precision of the density).

The $F_o - F_c$ (difference) map subtracts the observed and calculated density, giving rise to both positive and negative densities. Positive densities (typically shown by green in maps) implies that the contributions to the electron density from the observed data are larger than that from the current model, which indicates that the model does not contain enough electron density at that region. Negative densities (typically shown by red in maps) implies that there is too much density in current model than suggested by the experiment. At the end of the refinement, the $F_o - F_c$ map should be almost featureless and any significant features (> 3σ) indicate errors in the model. In practice the map coefficients are also weighted by maximum likelihood factors.[49]

Model validation: The quality of the final model can be evaluated by many indicators.[46] Here, I describe two important criteria used in my thesis (besides the difference density maps).

The R-factor is a global quality metric (i.e. it depends on the complete structure).[46] It monitors the agreement between the measured structure factor amplitudes $|F_o|$ and the amplitudes $|F_c|$ calculated from the current model. The R-factor is expressed as

$$R = \frac{\sum||F_o|-|F_c||}{\sum|F_o|} \tag{2.39}$$

If the observed and calculated amplitudes agree perfectly, $R = 0$. However, protein structures typically give R values between 0.1 and 0.3, and they depend also on the resolution of the data (lower values for high resolutions). Normally two types of R-factors are given for each structure. One is R_{work}, which is based on the entire reflection data and measures how well the current model can predict the entire set of data. The other is the free R-factor, R_{free}.[50] It is calculated based on only a small fraction of the reflections (5–10%), which is randomly selected and not included in the refinement. It is used to make sure that the model is not overfitted. The difference between R_{free} and R_{work} is considered as a measure of the overfit and should be kept minimal. Adding more parameters to the model (e.g. anisotropic instead of isotropic ADPs) always improves R_{work} but if R_{free} is not also significantly improved, it is only an effect of overfitting (fitting noise).

The real-space Z-difference (RSZD) score[51] is a local quality metric, used to judge whether a specific region of interest is correctly modelled (small changes in the structure have minimal influence on the R factors). It is based on the $F_o - F_c$ difference map and essentially indicates the largest positive or negative value in this map close to a certain atom or residue. Therefore, RSZD should be less than 3 for a good fit.

2.4.1. Crystallographic Refinement

Crystallographic refinement optimises the current model to improve the agreement with the observed reflection intensities.[46] This is done automatically by optimising an energy function of the form

$$E_{\mathrm{cryst}} = w_{\mathrm{A}} E_{\mathrm{Xray}} + E_{\mathrm{MM}} \tag{2.40}$$

where E_{Xray} is a penalty function that describes how well the model agrees with the experimental data. In principle, it could be any function that describes the difference between the current model and the experimental data, e.g. the R factor.

However, in practice, more sophisticated maximum-likelihood refinement target functions are used.[52,53] E_{MM} is an empirical restraint function that contain terms for bond, angle, dihedral and non-bonded restraints.[54] In terms of computational chemistry, it is an MM energy function, although it is often based on statistical evaluation of a large number of crystal structures, rather than being energy based, as normal MM functions. w_A is a weight factor, which is necessary because E_{MM} and E_{Xray} do not have the same units. It determines the relative importance of the crystallographic raw data and the MM force field for the final structure. The reason why the MM term is employed is that the crystallographic data is normally not accurate enough to give the exact position of all atoms. Therefore, the MM term is used to ensure that the final structure makes chemical sense, i.e. that it has proper bond lengths and angles and that conjugated groups are planar. Moreover, the non-bonded term ensures that atoms do not come too close. The refinement takes the form of a minimisation or simulated annealing calculation by MD. Today, refinement could be performed in many crystallographic software packages, e.g. Phenix, CNS, Refmac and SHELX,[55–57] which greatly eliminate the manual labour of the refinement.

2.4.2. Quantum Refinement

Quantum refinement is a method that combines crystallographic refinement and QM/MM calculations. As described above, standard crystallographic refinement takes chemical information of the protein into account by adding restraints, represented by MM force field. In quantum refinement, the MM restraints for a small region of particular interest in the protein are replaced by more accurate QM calculations.[58,59] This is especially important when we deal with unusual ligands (e.g. substrates or inhibitors) and metal sites, for which no geometry restraints or MM force field is available. Quantum refinement is described by the energy function

$$E_{Cqx} = w_A E_{Xray} + E_{MM12} + w_{QM} E_{QM1} - E_{MM1} \qquad (2.41)$$

Compared to Eqn. 2.40, quantum chemistry is introduced by replacing the MM potential for a small region of the protein (called system 1) by a QM calculation (in analogy to QM/MM calculations), yielding a QM energy for system 1, E_{QM1}. To avoid double counting, the MM energy of system, E_{MM1} must be subtracted from the MM energy of the entire protein, E_{MM12}. Thereby, we introduce an accurate energy function for the system of interest, which can greatly improve the geometries. The factor w_{QM} is another weight that is needed because the crystallographic MM force field is based on a statistical analysis of crystal structures.[60] Therefore, the force constants are not energy-derived, as is the QM term, and are in arbitrary statistical units. Besides this factor, Eqn. 2.41 is

obtained by replacing E_{MM} in Eqn. 2.40 by the QM/MM energy function in Eqn. 2.29.

Such an energy function is implemented in the software ComQum-X,[59] which combines the software Turbomole[61], for the QM calculations, and the crystallography and NMR system (CNS) software [62,63] for the crystallographic calculations. Minimal changes in CNS input files are made to write out data (energies, forces and coordinates) that can be combined by the corresponding QM data. The geometry optimisation of the QM system is performed by the optimiser in Turbomole.

Crystallographic refinement is traditionally performed without any electrostatic interactions, because hydrogen atoms are not discerned in the structure and hydrogen atoms are key to electrostatics interactions, e.g. in hydrogen bonds. We have followed this custom and excluded electrostatics and hydrogen atoms from all crystallography and MM calculations (but they are of course included in the QM calculations). In analogy with the QM/MM calculations, the QM system is truncated by H atoms, employing the link-atom approach.[35,36]

3. Nitrogenase

Nitrogen is an essential element, contained in most biomolecules essential to life, including all amino acids and the nucleotide components of DNA and RNA. Although this element is abundant in the atmosphere (78%) in the form of molecular nitrogen gas (N_2), nitrogen is often a limiting element for organisms. The reason for this is that the triple bond in N_2 is very strong, making it highly inert and hard to be metabolised. Therefore, it can only be used if it is in a reduced or oxidised form, like ammonia (NH_3) or nitrate (NO_3^-). The process of converting molecular nitrogen in the air into NH_3 or related nitrogenous compounds, is called nitrogen fixation.[3,64] Nitrogen fixation occurs naturally in the air by lightning. However, only a rather small contribution to the fixed nitrogen is supplied through this process.

Industrially, the most common method to produce NH_3 is the Haber–Bosch process, which requires high temperature, high pressure and an iron catalyst.[64] The process was invented in 1909–1910 and it is one of the most important industrial processes today, consuming 1–2% of the world's energy supplies and providing half of the total biologically available nitrogen on earth.[64] Thus, it is a major factor in the recent agricultural revolution and the human population explosion.[3]

Biologically, the enzyme nitrogenase (EC 1.18/19.6.1) is the only family of enzymes that can catalyse the reduction of N_2 to NH_3.[1–3] Given that about half of the fixed nitrogen sustaining Earth's organisms is formed through biological N_2 fixation, there is a great interest in understanding how the enzyme nitrogenase accomplish this task.[3] Nitrogenase is present only in a few groups of bacteria and archaea.[1–3] However, several of these live in symbiosis with higher plants, such as legumes, rice, sugarcane and alder. The nitrogenase reaction is performed at ambient temperature and pressure. Still, the process is very energy demanding, consuming 16 molecules of ATP for each molecule of N_2 fixed, according to the reaction:

$$N_2 + 8\ e^- + 8\ H^+ + 16\ ATP \rightarrow 2\ NH_3 + H_2 + 16\ ADP + 16\ P_i \qquad (3.1)$$

from which it can be seen that H_2 is also produced by the reaction.

Nitrogenases have been extensively studied by experimental methods.[1–3,65,66] Several crystal structures of nitrogenases have been determined,[5,48,67–70] showing that the enzyme is comprised of two protein components. One is the MoFe protein (Figure 3.1a), a $\alpha_2\beta_2$ heterotetramer that contains two iron–sulfur clusters: the $MoFe_7CS_9$ homocitrate iron–molybdenum (FeMo) cluster, believed to be the active site for substrate binding and reduction (Figure 3.2a), and the Fe_8S_7, P-cluster (Figure 3.2b), presumed to mediate electron transfer. The other one is the Fe protein (depicted in Figure 3.1b), a homodimer that contains a Fe_4S_4 cluster and binding sites for two ATP molecules. The reduced Fe protein binds two molecules of ATP which in turn change its conformation, triggering the docking to the MoFe protein and the intramolecular and intermolecular electron transfer through Fe protein to FeMo cluster, via the P-cluster. Hydrolysis of ATP molecules triggers the disassociation of the Fe protein from MoFe protein. Then, a new reduced and ATP-loaded Fe protein may bind in next catalytic cycle. Molybdenum-dependent nitrogenase is the most widely studied enzyme, but in some variants of the enzyme, the Mo ion in FeMo cluster is replaced by vanadium or iron,[71] denoted V- or Fe-type nitrogenases. The Mo nitrogenase is the most active and has been studied extensively.[3]

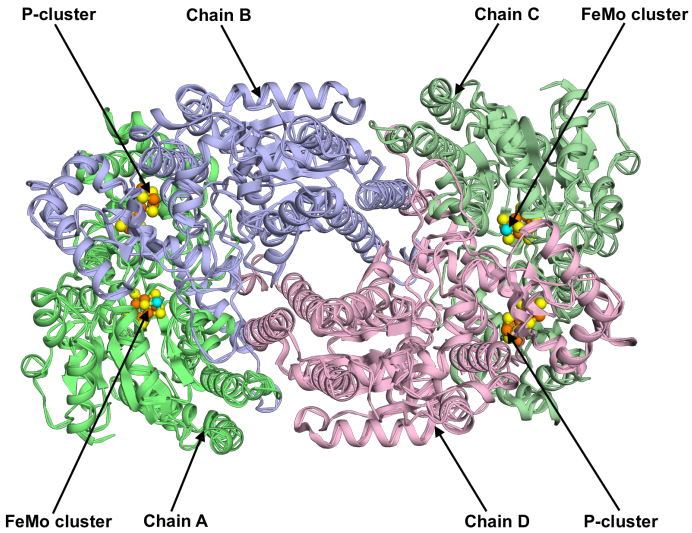
3.1. FeMo Cluster

The FeMo cluster is believed to be the active site of nitrogenase, where dinitrogen is converted to bioavailable ammonia during the accumulation of eight electrons and protons, together with the obligatory formation of a H_2 molecule.

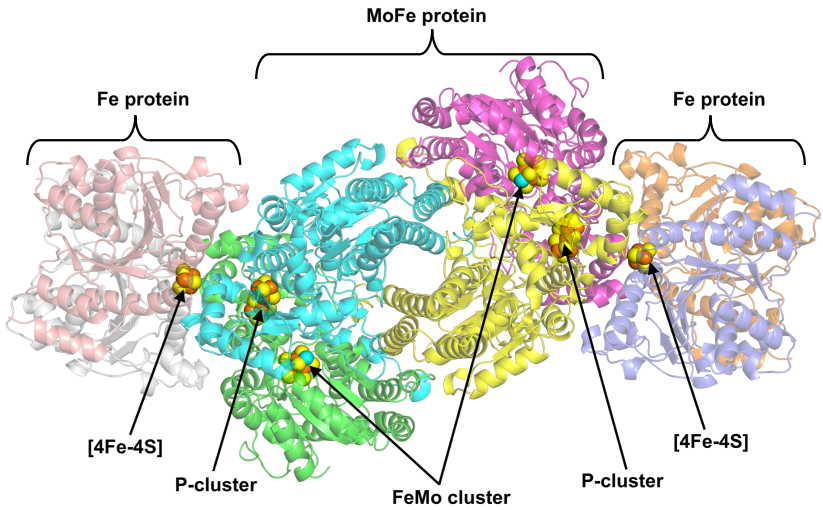
3.1.1. Atomic Structure

A knowledge of the detailed atomic structure of the active site is fundamental to understand the chemistry of the complex enzyme. Owing to intensive work over the past years, remarkable structure details have been provided.[5,48,67–70] The first X-ray crystallographic structure of the MoFe protein of Mo-type nitrogenase from *Azotobacter vinelandii*, was determined by Rees and Kim in 1992 at 2.7 Å resolution.[68] In this structure, the FeMo cluster was described as $MoFe_7S_9$(homocitrate) with a central cavity surrounded by Fe and S ions. In 2002, a higher-resolution crystal structure (1.16 Å) was presented, which surprisingly showed a light atom, coordinated to six irons in the centre of the FeMo cluster.[69] The identity of this atom (C, N or O) was long debated and many scientists assumed it was a N^{3-} ion. Almost 10 years later, Rees, Einsle and coworkers clarified that the previous assignment was not correct. Integrating electron density at the cluster centre, together with electron spin echo envelope

modulation (ESEEM) spectroscopies supplied evidence that the interstitial atom is C rather than N, based on an improved crystal structure at 1.0 Å resolution.[48] Simultaneously, DeBeer and coworkers performed X-ray emission spectroscopy (XES) studies and demonstrated that the presence of C atom fits the XES data best, confirmed also by theoretical studies.[65]



a



b

Figure 3.1: a) The nitrogenase MoFe protein from *Azotobacter vinelandii* (PDB code 3U7Q).[48] The structure is a dimer of heterodimers and the four subunits are shown in different colours. b) The complex between the Fe and MoFe proteins of nitrogenase (protein data bank (PDB) code: 1N2C).[72] The metal clusters are highlighted in a space-filling model.

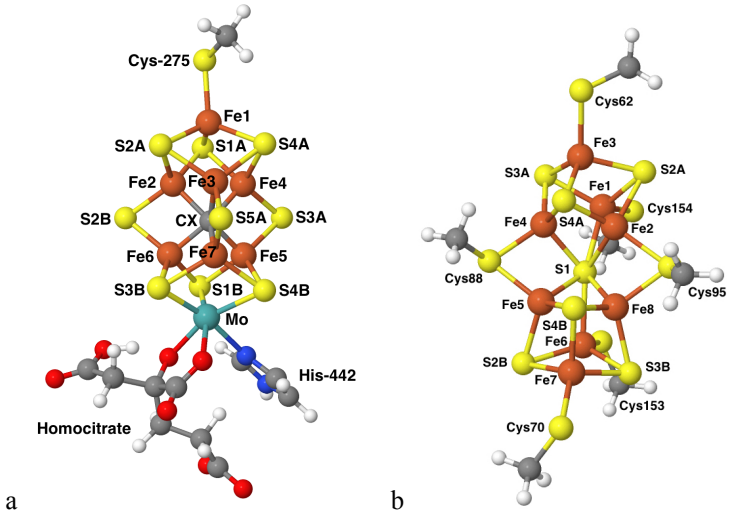


Figure 3.2: The atomic structure of a) FeMo the cluster and b) the P-cluster with atom and residue names from the 3U7Q crystal structure.[48]

Thereby, the basic molecular structure of the FeMo cluster is known, displayed in Figure 3.2a. It is a $MoFe_7CS_9$(homocitrate) cluster, connected to the enzyme by one cysteine (Cys) and one histidine (His) residue.

3.1.2. Redox Properties

Another fundamental property of the cluster, which is important for the understanding of its function and necessary for its computational modelling is the oxidation-state assignment of the metal ions and hence, the total charge of the cluster. The cluster has been extensively characterised by many experimental techniques such as electron paramagnetic resonance (EPR) spectroscopy, Mössbauer spectroscopy, enhanced nuclear double-resonance (ENDOR) spectroscopy and X-ray absorption spectroscopy (XAS). They have shown that the FeMo cluster is in a $S = 3/2$ spin state in the resting state, with three unpaired electrons.[3,73,74] The sulfide (S^{2-}) and carbide ions (C^{4-}) are normally assumed to be in their usual closed-shell –II and –IV oxidation states, respectively. The oxidation state of the molybdenum atom has been widely studied since the 1970s.[15–20] It was originally assumed to be in the +IV oxidation state, i.e. the reduced state of most other Mo enzymes and three oxidation state-models were discussed: [$2Fe^{2+}$:$4Fe^{3+}$:Mo^{4+}], [$4Fe^{2+}$:$2Fe^{3+}$:Mo^{4+}] and [$6Fe^{2+}$:$1Fe^{3+}$:Mo^{4+}].[78–80] However in 2014, it was shown that the Mo ion is be better described as Mo^{3+} with a non-Hund (↑ ↓ ↓) electron configuration, identified by a combined experimental (XAS) and computational (time-dependent density functional

theory, TD-DFT) approach.[66] Thus, the three oxidation state models were adjusted to [$1Fe^{2+}$:$6Fe^{3+}$:Mo^{3+}], [$3Fe^{2+}$:$4Fe^{3+}$:Mo^{3+}] and [$5Fe^{2+}$:$2Fe^{3+}$:Mo^{3+}]. Soon afterwards, combined XAS and computational studies settled that three of the seven irons in the cluster are in the reduced state (Fe^{2+}) with the remaining four irons in the oxidised state (Fe^{3+}).[81] This settled the oxidation-state distribution, giving a total charge of –1 for the $MoFe_7CS_9$ core in the resting state.[82]

3.1.3. Reaction Mechanism

Much experimental evidence has identified the FeMo cluster as the place where the catalytic chemistry of nitrogenase occurs.[83] Even if the atomic and the electronic structure of the FeMo cluster is clarified, the detailed catalytic reaction mechanism of nitrogenase remains elusive. Based on kinetic studies, Thorneley and Lowe suggested that the N_2 activation and reduction to NH_3 can be described by a cycle involving nine catalytic intermediates E_0–E_8 differing in the number of electrons and protons delivered to the FeMo cluster.[84] This is called the Lowe–Thorneley kinetic scheme. A simplified model is displayed in Figure 3.3, showing that N_2 cannot bind to the cluster until either the E_3 or E_4 state, i.e. after three or four electrons and protons have been added to the resting state E_0. Most subsequent experimental and theoretical investigations have been based on this scheme. Recently, several of these intermediates have been trapped and spectroscopically characterised.[3] Of particular interest is E_4, the Janus intermediate, which is believed to be the species that binds N_2.

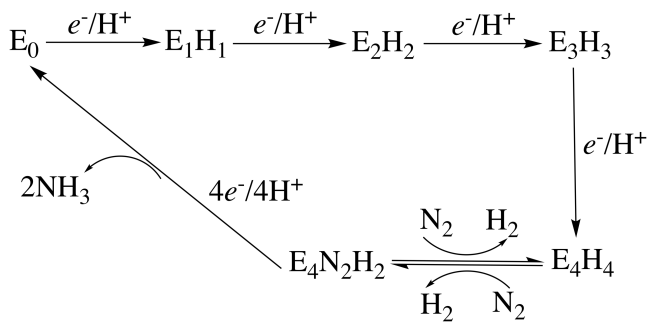


Figure 3.3: A simplified Lowe–Thorneley scheme.[3,84]

Despite extensive experimental investigations, the precise binding position for the substrate remains largely unknown. A large set of experimental MoFe protein residue mutations have been performed. These have shown that substitution of the α-subunit His-195 by glutamine (Gln) gives an inactive enzyme, indicating

that it is necessary for N_2 reduction.[85,86] Moreover, Dean and coworkers have constructed mutations of α-subunit valine-70, leading to the conclusion that this residue imposes steric constraints for substrates accessing one of the three 4Fe4S faces, viz. the one involving Fe2, Fe3, Fe6 and Fe7 (the names of the Fe ions are taken from the crystal structure 3U7Q,[48] shown in Figure 3.2a).[83]. Consequently, most studies have assumed that this is the reactive face of the FeMo cluster.

Intermediates formed during the reaction are very short-lived, making it hard to trap them and even harder to characterise them. Only recently, E_1 was elucidated, showing that it contains a protonated sulfide ion that was identified as S2B, based on a computational study.[87] Likewise, the E_4 state has been characterised by recent EPR and freeze-trapping experiments, significantly advancing our knowledge about the reaction mechanism.[1–3,65,66] The freeze-trapped E_4 intermediate, which is the central state in the Thorneley–Lowe scheme, should contain four protons. Hoffman and coworkers have proposed, based on the interpretation of high-resolution ENDOR and EPR spectroscopies that two of the protons are sulfide bond protons, whereas the other two are hydride bridges between two pairs of Fe ions.[3,88,89] With the help of DFT calculations, they derived a structure in which the two hydrides and the two protons bind on the same face of the cluster, viz, between the Fe2/6 and Fe3/7 pairs, as well as on the belt S2B and S5A ions, as shown in Figure 3.4. Moreover, they suggested that the binding of N_2 and the first two protonations of N_2 are accompanied by the reductive elimination of the two hydride bridges as a H_2 molecule, which provides the thermodynamic driving force and facilitated the binding of N_2, making the reaction thermoneutral.[3,90–92] This provides an attractive explanation to the compulsory formation of H_2.

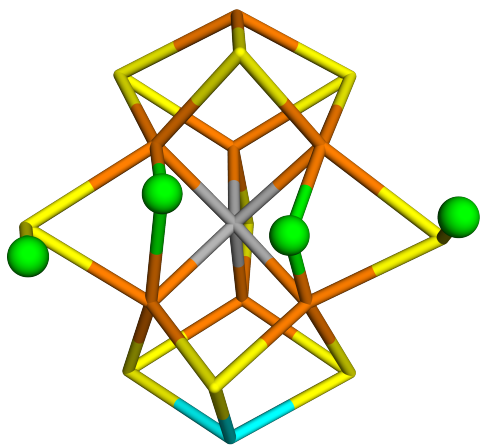


Figure 3.4: The structure of the E_4 state, proposed by Hoffman and coworkers[3,88,89] as obtained by our QM/MM geometry optimisations. The protons (show as green balls) bind to the S2B and S5A ions, whereas the hydride ions (also green balls) bridge the Fe2/6 and Fe3/7 pairs.

After the binding of N_2, two reaction schemes for the formation of NH_3 have been discussed. The first is the distal pathway, which was originally suggested by Chatt.[93,94] In this mechanism, one N atom of N_2 is first protonated, so that the first NH_3 leaves at the E_5 stage, before the other N atom is protonated. It has gained support from inorganic model complexes.[95–97] However, it has been suggested to apply also for nitrogenase by several authors.[98,99] In the second mechanism, the protons are added alternatively to both N atoms, so that HNNH and H_2NNH_2 are intermediates and the first NH_3 molecule does not dissociate until the E_7 stage.[3,100] This alternating pathway is supported by the fact that nitrogenase can use hydrazine as a substrate and that hydrazine is released upon acid or base hydrolysis of the enzyme during turnover.[1,3,101,102] Moreover, it has been shown that N_2, N_2H_2, CH_3N_2H and N_2H_4 all react via common intermediate and therefore share a common mechanism.[3,100] The two mechanisms are illustrated in Figure 3.5.

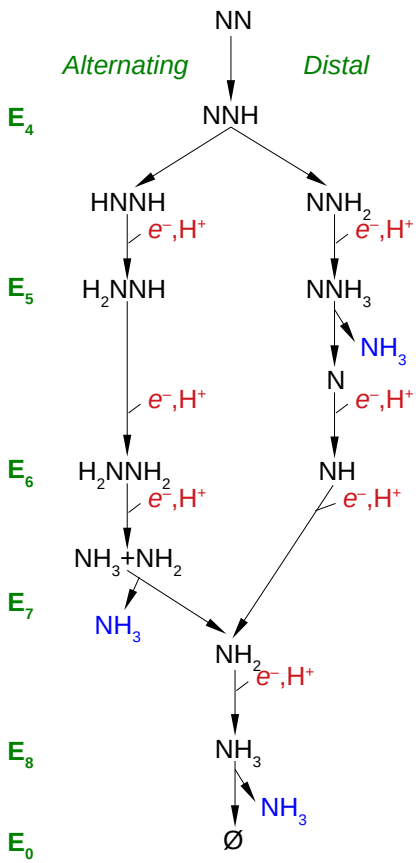


Figure 3.5: The difference between the alternating and distal mechanism of the later part of the Lowe-Thorneley reaction scheme of nitrogenase.[3]

3.1.4. Previous Computational Studies

Nitrogenase has also been extensively studied by computational methods.[3,66,98,103–114] Earlier studies were hampered by the fact that the central carbide ion in the FeMo cluster was not identified until 2014[5,48,65,69,70] and that the redox state of the cluster was not settled until 2017.[81,82,115] Moreover, many earlier studies did not involve the loading of the active site by four electrons and protons before the actual reaction could take place. However, even among the latest QM studies, there are no agreement among the reaction mechanism.

For example, Nørskov and coworkers have proposed that dissociation of a doubly protonated S2B and the evolution of H_2 precede the N_2 reduction process, promoting the N_2 activation on the exposed Fe sites of FeMo cluster.[98] Moreover, they suggested that N_2 binds end-on to the cluster, with one N atom bridging two Fe ions. The first NH_3 is formed via subsequent protonation steps of the distal N atom and it dissociates at the E_5 level, before the second N atom starts to be protonated (i.e. a distal mechanism). The second NH_3 molecule is formed at the E_8 level, it dissociates and S2B binds again, completing the catalytic cycle.

On the other hand, Dance has proposed a mechanism wherein the E_4 state involves hydrogen atoms on Fe6, S2B and Fe2. The fourth proton is on S3B and he suggests that all protons are supplied from the solvent through this sulfide ion. In contrast to the end-on binding mode in the Nørskov mechanism,[98] Dance suggests that N_2 binds side-on to the Fe6 and Fe2 ions and is protonated alternatively on the two N atoms.[104,116] More recently, he compared 11, 31, 33 and 35 structures for E_1, E_2, E_3 and E_4 states and concluded the most stable structure for each state has protons on S2B, S2B–Fe6, S2B–Fe6–Fe2/6 and S2B–Fe6–Fe2–Fe6 respectively.[116]

Siegbahn has argued that it is energetically much more favourable if the protons bind to the central carbide ion than as hydride ions binding to the Fe ions. Thus, he has suggested that the most stable E_4 structure should have one proton on a sulfide ion and three protons on the carbide ion, forming a CH_3^- group, which moves out from the centre of the cluster, forming a cavity where N_2 can subsequently bind and react.[105,117,118] In addition, he suggested that N_2 binds side-on, bridging between two Fe ions after the cluster is reduced by in total six or eight electrons (i.e. that the resting E_0 state is not involved in the reaction cycle).[117–119]

McKee made DFT calculations on an extremely simplified model, with no ligands or surrounding residues included and with Fe instead of the Mo ion. Thus, the model included only 18 atoms.[106] In his proposed mechanism, the first four protons bind to S2B, S3B, the central carbide and as a hydride ion bridging between Fe6 and Fe7. However, the H_2 elimination takes place from a metastable

state, in which the H atom on carbide ion shifts to bridging between Fe2 and Fe6. After the elimination of H_2 (from the two bridging hydride ions), N_2 binds terminally to the central carbide ion. The reaction then proceeds by a distal mechanism in which one N atom is first triply protonated until NH_3 is formed and released. Subsequently, the other N atom is protonated.

Adamo, Xu and Rao have suggested that the most stable E_4 state has protons on S2B, S2A, S5A and the central carbide ion.[107] In the next step, N_2 binds covalently to the central carbide ion, whereas the protons on S2A and C forms a H_2 molecule that dissociates from the cluster. Their suggested mechanism involves the following intermediates: C–NNH, C–NNH_2, C–$NHNH_2$, C–NH_2NH_2, release of the distal NH_3, C–NH_2 and C–NH_3 (i.e. all reaction intermediates are covalently bound to the central carbide ion).

Additional computational studies of nitrogenase have been presented,[4] but already from this short survey, it should be clear that there is not at all any computational consensus regarding the reaction mechanism of nitrogenase.

3.1.5. Broken-Symmetry States

The FeMo cluster comprises eight transition metals. Therefore, the electronic structure is complicated. In the resting state, it is known from EPR and Mössbauer spectroscopy that it is in a quartet spin state, S = 3/2, with three unpaired electrons. However, this state arises from the a coupling of high-spin Fe^{2+} and Fe^{3+} ions (with four or five unpaired electrons) and the unusual non-Hund state of Mo^{3+}.[3,73,74] Such a complicated electronic structure requires an advanced multiconfigurational QM description.[120,121] Unfortunately, such studies are currently very demanding. Therefore, the great majority of computational studies of nitrogenase have employed density functional theory (DFT), using the broken-symmetry (BS) approach developed by Noodleman and coworkers.[122] This approach employs separate wave functions for the electrons with α and β spin. In practice, this means that each of the seven Fe ions are modelled in their high-spin state with either a surplus α (four Fe ions) or β (three Fe ions) spin and that they then couple antiferromagnetically to a lower net spin state.

In 2001, Noodleman and coworkers presented a detailed DFT study of the FeMo cluster with C_3 symmetry, showing that 10 BS states can be obtained for all the possible arrangements of α and β spins on the seven Fe irons.[112] These ten BS states are schematically depicted in Figure 3.6. However, recognition of the actual C_1 symmetry of the cluster yields 35 BS states ($7!/(3! * 4!)$).[123] In their first study, Noodleman and coworkers suggested that BS6 has the lowest energy and that the properties of this state fitted the X-ray structure and experimentally observed hyperfine spectra from Mössbauer and ENDOR spectroscopies better than other tested BS states.[112] However, these calculations were performed

without any central atom in the FeMo cluster. In 2007, they extended the study to models with a central atom (N, C or O) and found that BS7 state was 7 kcal/mol lower in energy than BS6.[124]

BS7 was also found to be the ground state in calculations with an appreciably larger (225-atom) QM-cluster model[66] and it has been used in most recent QM studies of nitrogenase,[65,66,78,115,124,125] although some studies instead used the BS6[104,113] or BS2 states.[105] In calculations that automatically find the lowest BS state, either BS6 or BS7 was found to be the ground state, depending on the oxidation state of the cluster and the bound ligands (substrate and protons) during the reaction mechanism.[114]

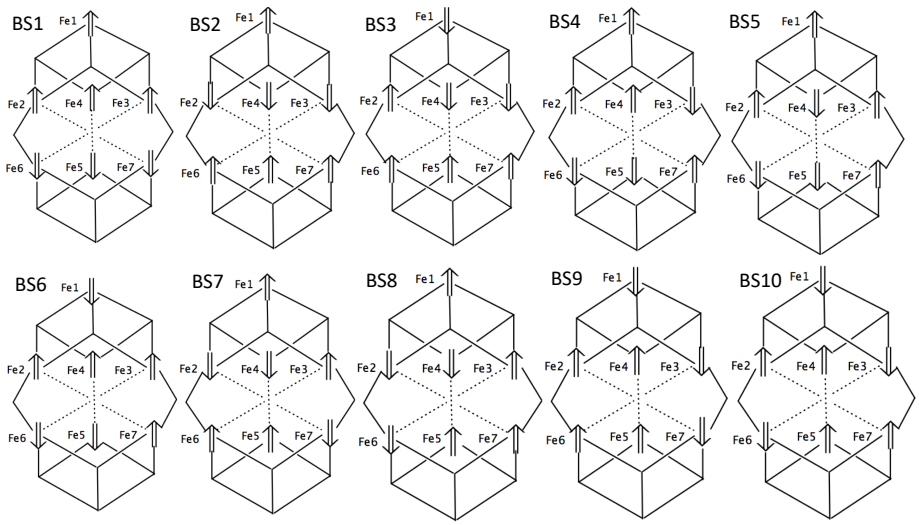


Figure 3.6: Nodlemann's ten BS states of the FeMo cluster in nitrogenase.[112]

For the spin state, Siegbahn suggested that the resting state modelled by a doublet BS2 state was lower in energy than that of a quartet state, which is in conflict with the experimental finding of a quartet state.[105] He used the singlet and doublet states for all other intermediates in the reaction mechanism (for which no experimental information is available), whereas Kästner and Blöchl reported higher spin states (triplets–sextets) for all their intermediates.[114]

In our calculations, we have thoroughly studied the energies of the various BS states. We usually obtained a starting wave function by first optimising the all-high-spin (HS) state, with all 35 unpaired electrons aligned in the same direction. Then, we changed the α and β occupation number to get a desired total spin state.[112] This typically gave one BS state. Other BS states were obtained by simply swapping the coordinates of the Fe ions.[126] In some cases, we instead use fragment approach by Szilagyi and Winslow to obtain a proper starting state.[127]

3.2. P-cluster

3.2.1. Atomic Structure and Redox Properties

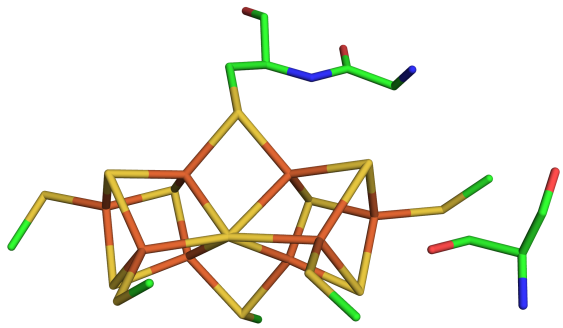
As already mentioned, nitrogenase contains another iron–sulfur cluster besides the FeMo cluster, viz. the P-cluster. It is localised in the MoFe protein at the interface between the two subunits. Crystallographic studies have shown that it is a $Fe_8S_7Cys_6$ cluster.[5,48,67] In the reduced resting state, it is essentially composed of two [4Fe-4S] cubane clusters, which share one sulfide ion (called S1; the naming of atoms and protein residues is taken from crystal structure of nitrogenase from *Azotobacter vinelandii*[48]), which coordinates to six Fe ions (Figure 3.7a). Two of the Cys ligands also bridge two pairs of Fe ions from two cubane subunits. Thereby, all the eight Fe ions have three sulfide ligands and one Cys ligand each.

Interestingly, the resting state of the P-cluster in nitrogenase is in the fully reduced $Fe(II)_8$ state (P^N). This is in sharp contrast to the [2Fe-2S] ferredoxins, which employ the Fe(II)Fe(III) and $Fe(III)_2$ states, and the [4Fe-4S] ferredoxins and high-potential iron proteins, which employ the $Fe(II)_3Fe(III)$, $Fe(II)_2Fe(III)_2$ and $Fe(II)Fe(III)_3$ states.[128] The P-cluster also employs the one-electron oxidised state (P^{1+}) and perhaps also the two-electron oxidised state (P^{2+}), whereas the three-electron oxidised state (P^{3+}) is most likely not involved in catalysis.[129–131] More oxidised states have also been observed, but they are irreversible.[130]

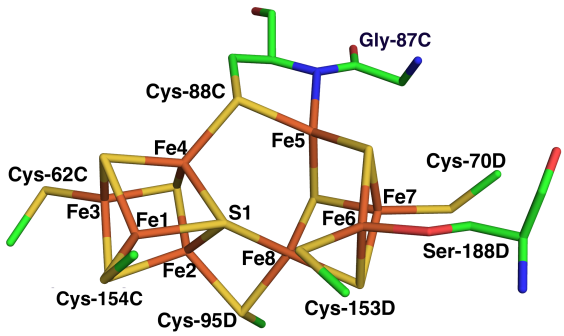
The oxidised states show conspicuous changes in the geometry of the P-cluster. In P^{2+}, for which an accurate crystal structure exists (Figure 3.7b),[48] the backbone N atom of Cys-C88 coordinates to Fe5 and the side-chain O atom of Ser-D188 coordinates to Fe6. This is accomplished by 1.3–1.4 Å movements of these two Fe ions, whereas all the other atoms in the P-cluster essentially remain in the same position. Thereby, the S1–Fe5 and S1–Fe6 bonds are cleaved and are replaced by the new Fe5–N and Fe6–O bonds. Consequently, all the Fe ions in the P-cluster remain four-coordinated, but one corner in the Fe5–Fe6–Fe7–Fe8 subcluster is lost and the structure becomes more distorted. It is notable that Ser-D188 is not strictly conserved among nitrogenases, but those that miss this residue has a tyrosine at position 189 that may also coordinate to the P-cluster.[132]

The structure of the P^{1+} state has been more controversial, but in 2018 a crystal structure of nitrogenase was published that was suggested to show the P-cluster in the P^{1+} state.[47] The structure was intermediate between the P^N and the P^{2+} states: It contains the Fe6–O bond with Ser-D188, but not the Fe5–N bond. Instead, Fe5 still binds to S1. Thereby, the second cubane is somewhat less distorted, as can be seen in Figure 3.7c.

a



b



c

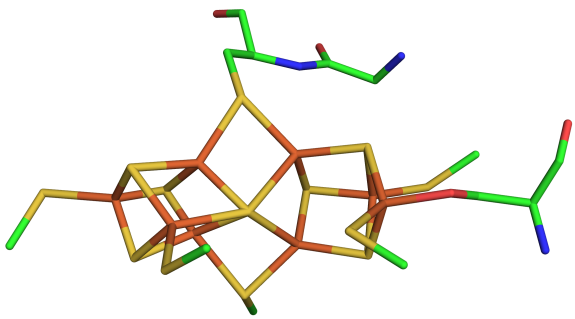


Figure 3.7: Redox-dependent conformational changes of the P-cluster. Shown are the (a) P^N resting state, (b) the P^{2+} doubly oxidised state and c) the P^{1+} partly oxidised state. In b) the residue and atom names are indicated. a) and b) are derived from the 1.0 Å-resolution crystal structure of the nitrogenase from *Azotobacter vinelandii* (PDB code 3U7Q[48]), whereas c) comes from the 2.1-Å structure from the same organism (PDB code 6CDK)[47].

The coordination of the backbone N atom of Cys-C88 clearly requires that this atom is deprotonated (otherwise, the proton would interfere with the binding). For the side-chain O atom of Ser-D188, the situation is less clear: Both a protonated and a deprotonated alcohol can coordinate to Fe, but it is expected that the pK_a of the alcohol is significantly lowered. It is also possible that the two groups are deprotonated both when coordinated and when it does not bind to the

metal. Experimentally, it has been observed that that the P^{2+}/P^{1+} conversion is pH-dependent.[133,134] However no pH dependence was found for the P^{1+}/P^{N} couple. This has complicated the interpretation of the structures of the P-cluster.

3.2.2. Electron-Transfer Reactions

The function of the P-cluster is to mediate electrons between the [4Fe-4S] cluster in the Fe protein and the catalytic FeMo cluster in the MoFe protein. In the crystal structure of the complex of the Fe and MoFe proteins,[72] the P-cluster is located in between the other two clusters, with a distance of ~14 Å to both (Figure 3.1b).

Several redox potentials of the P-cluster have been measured. Interestingly, the P^{1+}/P^{N} and P^{2+}/P^{1+} potentials are identical, –309 mV.[2] However, the potential is significantly lowered when the Fe protein docks, especially when loaded with ATP, P^{2+}/P^{1+} to –430 mV and P^{1+}/P^{N} to below –550 mV.[2]

The electron transfer in the nitrogenase complex has been suggested to follow a deficit–spending model,[135] in which the electron is first delivered from the P-cluster to the FeMo cluster in a slow step. Then, the P-cluster is rapidly reduced by the cubane in the Fe protein. The Fe protein docks to the MoFe protein only when reduced and loaded by two molecules of ATP. It is the only protein that can deliver electrons to the MoFe protein. After electron delivery, the two ATP molecules are hydrolysed to ADP, after which the Fe protein dissociates. The latter dissociation is the rate-limiting step of the nitrogenase reaction. This process has to be repeated eight times during the reduction of one N_2 molecule to two molecules of ammonia.

In such a mechanism, only the P^{N} and P^{1+} oxidation states of the P-cluster are involved. It is then unclear how a single redox couple of the P-cluster with a constant redox potential can provide electrons to the FeMo cluster in all eight E_0–E_7 states, with a large variation of the redox level and bound ligands. An attractive explanation is that also the FeMo cluster alternates between only two redox levels.[3,136] When the first electron and proton are taken up during the $E_0 \rightarrow E_1$ transition, the cluster is reduced from the $Fe(II)_3Fe(III)_4$ state to the $Fe(II)_4Fe(III)_3$ state and the proton binds to a sulfide ion (Mo always remains in the Mo(III) state). However, when the next electron and proton are delivered during the $E_1 \rightarrow E_2$ transition, the proton binds to one or two Fe ions (probably the Fe2/6 pair) in the form of a hydride ion. This means that it takes up two electrons, so that the FeMo cluster in the E_2 state is back to the $Fe(II)_3Fe(III)_4$ redox level. The same procedure is repeated in the $E_2 \rightarrow E_3 \rightarrow E_4$ transitions: The first proton binds to a sulfide ion, whereas the second binds as a bridging hydride ion, giving $Fe(II)_3Fe(III)_4 \rightarrow Fe(II)_4Fe(III)_3 \rightarrow Fe(II)_3Fe(III)_4$ transitions, i.e. oscillations forth and back between only two redox states.

However, what is then the relevance of the P^{2+} state, which is so stable that it is observed in crystal structures?[48] Owens et al. have shown that the coordination of the P-cluster by a backbone amide N atom and a hard O-based amino acid ligand is highly conserved among nitrogenases from different organisms.[132] Rupink and coworkers have proposed that the P-cluster may act as two coupled $[Fe_4S_4]$ clusters, each capable of donating one electron to the FeMo cluster.[131] This might be relevant in later stages of the nitrogenase reaction.

3.2.3. Electronic Structure

Like the FeMo cluster, P-cluster has a complicated electronic structure. The individual Fe ions each are in the high-spin Fe(II) or Fe(III) state, but they then couple antiferromagnetically to a lower net spin state. The two cubane subclusters are often considered to couple independently.[137] As for the FeMo cluster, such electronic structures are in DFT treated by the broken-symmetry approach (BS). Chan and coworkers have presented more correct density matrix renormalisation group complete active-space calculations, showing a much larger number of possible states.[138]

Experimentally, the net spin state of the P-cluster is known for the four most reduced redox states:[137,139] P^N is a diamagnetic singlet ($S = 0$), P^{2+} has either $S = 3$ or 4 (Mouesca et al. has argued for the latter state[137]), whereas the P^{1+} is a physical mixture of $S = 1/2$ and 5/2 states and P^{3+} is a physical mixture of $S = 1/2$ and 7/2 states. For the P^N state, Mouesca et al. suggested that each subcluster has a vanishing net spin, which can be obtained by two Fe ions with α spin and two with β spin. For each subcluster, such a state can be found in 6 different ways, giving 36 possible BS states for the full P-cluster. However, only 18 of these are distinct, because the selection of α and β spin is arbitrary. Such states are also possible for the $S = 1/2$ P^{1+} state and there all 36 BS states are distinct.

On the other hand, it has been observed that the fully reduced [4Fe-4S] clusters are in the $S = 4$ state.[140,141] Such a state can be obtained with three Fe ions with α spin and only one with β spin. Combining two such states with opposite spin would give the $S = 0$ state for the P-cluster and they give rise to $4 \times 4 = 16$ different BS states. The study by Chan and coworkers suggested such a state for the P-cluster.[138]

The states with higher values of S can be obtained by assuming that only one Fe ion has β spin in one subcluster and two in the other, although this requires that one of the Fe ions is in the intermediate spin state. Such BS states can be obtained in $2 \times 4 \times 6 = 48$ different ways. In Paper VIII, we investigated which of all these BS states is most stable for each of the four redox states of the P-cluster.

4. Summary of the Articles

This thesis is based on ten publications, Papers I–X. Below, I will summarise each of them, discussing the aim, the methods used and the main results. All papers are about nitrogenase, but Papers II–VII deal with the FeMo cluster, whereas Papers VIII and IX deal instead with the P-cluster. All publications employ QM/MM calculations, but Papers I, III and VIII–X involve quantum-refinement calculations as an important ingredient. In Paper IX, we even make a method development to extend this approach to treat proteins with disorder in the QM system. In Paper I, we use MD simulations, as well as several special QM free-energy methods[39,142,143] to decide the most stable protonation states of some key residues close to the FeMo cluster.

4.1. Paper I

Almost all previous QM studies on nitrogenase were performed with QM-cluster calculations, which ignore the surrounding protein or model it as a featureless continuum solvent.[4,66,105,107,112] We think that there is a large risk that such calculations are inaccurate, because they ignore the electrostatic influence from the protein and the cluster may allow a too large flexibility. Therefore, we chose to instead use the QM/MM approach, which includes the full protein and a significant layer of solvating water molecules in the calculations. However, such calculations are much more demanding to set up than QM-cluster calculations, because the entire protein is included in atomic detail. Knowing that all the subsequent calculations would use this model and also that all previous QM studies have given conflicting results, we spent unusually much time and effort to set up the calculations.

First, we had to decide how much of this large protein should be included in the model. A direct look at the crystal structure (Figure 3.1a) shows that the four subunits are closely intertwined (for example, the P-cluster is at the interface between subunits A and B or C and D), indicating that there is no natural way to reduce the model. Therefore, we decided to include the complete tetramer in the model. Thereby, we also get a more realistic model (a truncated model will always replace some parts of the protein with solvent).

Second, we needed to settle the protonation state of all residues in the protein, because the electrostatics dominates the intermolecular interactions in proteins (protonation changes he charge of the residue) and the protonation states also affect hydrogen bonds. Initial protonation states of all the residues were determined from a detailed study of the hydrogen-bond pattern and the solvent accessibility. It was checked by the PROPKA[144] and Maestro[145] software. For most of the 2000 residues in nitrogenase, the protonation state is evident from the crystal structure or the residue is so far away from the active site that it is unlikely to significantly affect the results. However, in a few cases, residues close to the FeMo cluster or the P-cluster had uncertain protonation states. Therefore, we studied the protonation states of eight protein residues, including four His residues (His-195, 274, 362 and 451), three Glu residues (Glu-153, 380 and 440) and one Asp residue (Asp-445). These residues are shown in Figure 4.1.

This was done by running MD simulations of different possible protonation states. Then, we tried to decide the best protonation state by comparing the hydrogen-bond pattern with heavy-atom distances in the crystal structure, as well as the root-mean-squared deviation (RMSD) of the various atoms from the starting crystal structure.[48] These calculations (especially the RMSD) gave quite clear and consistent results (cf. Figure 4.2): His-195 and 362 are protonated on the NE2 atom and not on the ND1 atom, His-274 and 451 are protonated on the

ND1 atom and not on NE2, Glu-153 is protonated, Glu-380 deprotonated, whereas Glu-440 is protonated with the HE2 atom directed towards the deprotonated Asp-445 residue.

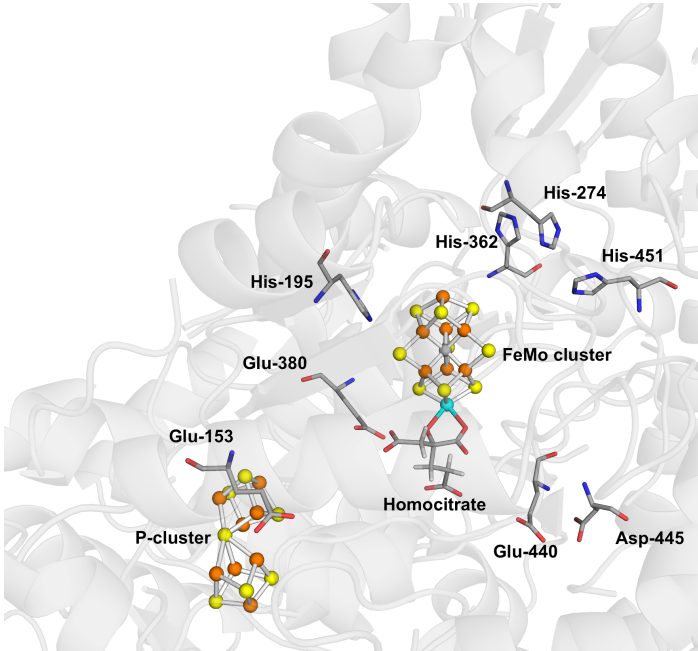


Figure 4.1: The nine residues around the FeMo cluster and the P-cluster in nitrogenase, for which we studied the protonation state in Paper I.

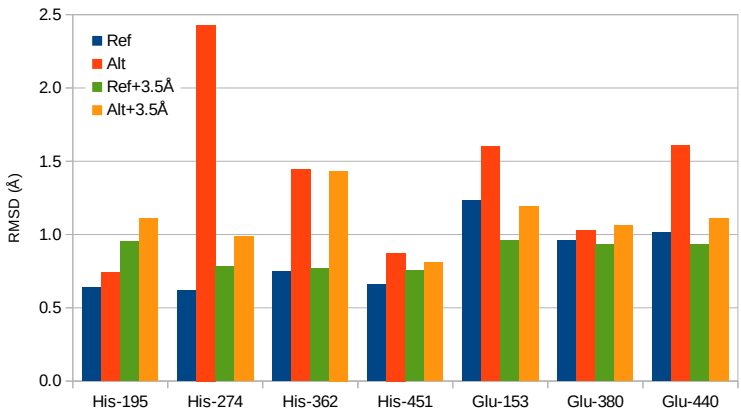


Figure 4.2: The results of the RMSD analysis, showing that the reference MD simulation (Ref., i.e with the selected protonation) always gives a lower RMSD to the starting crystal structure[48] than the MD simulations with each residue in the alternative protonation state (Alt). The RMSD analysis is done either only for the studied residue of for all residues within 3.5 Å of the studied residue.

Of particular interest is the homocitrate residue, which is a ligand of the Mo ion and therefore directly affects the electronic structure of the cluster and can also be involved in hydrogen bonds to the substrate. Homocitrate contains one alcohol and three carboxylate groups (cf. Figure 4.3a). In water solution at neutral pH, the alcohol is protonated and the three carboxylate groups are deprotonated. However, this might change when it coordinates to the Mo ion. It has been modelled in many different ways in previous QM studies, e.g. by smaller molecules [78,98,104,110] or by the entire homocitrate molecule with a charge of –2,[105] –3[66] or –4.[66,107,112,113]

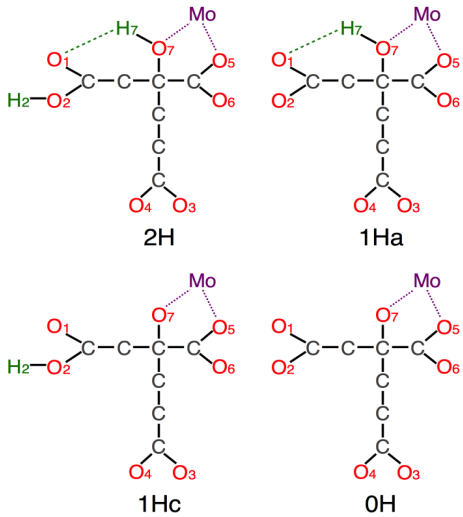
We first performed MD simulations and RMSD analysis for the homocitrate with four different protonation states, i.e. both O2 and O7 protonated (2H, net charge –2), both deprotonated (0H; net charge –4), or either O2 (1Hc) or O7 (1Ha) protonated ("c" and "a" indicate protonation of the carboxylate or alcohol; net charge –3). These states are shown in Figure 4.3a (other possible protonation states were discarded based on the interactions found the crystal structure). The MD simulations gave no clear conclusions from the H-bond analysis. However, the 1Ha state gave the lowest RMSD compared to the crystal structure.

We also performed several sets of QM calculations of the four protonation states at different levels of approximation, in order to predict the protonation state by analysing the p*K*a values and the relative energies. QM-cluster calculations in a continuum solvent pointed out 1Ha as the most stable state at pH 7, but the 2H state was only one pK_a unit less stable. QM/MM calculations also supported 1Ha as the most stable state, but QM/MM-PBSA (QM/MM combined with Poisson–Boltzmann and solvent-accessible surface area solvation[143]) gave more varying results, owing to the strong solvation effects. Still, the more accurate QTCP (QM/MM thermodynamic cycle perturbation) calculations[142,146] confirmed that the 1Ha state is most stable, four pK_a units more stable than the 2H state. Together, these calculations clearly showed how hard it is to estimate pK_a values of chemical groups inside a protein and close to metal ions.[147–149] In particular, continuum methods are problematic because the results strongly depend on the dielectric constant of the protein, which is unknown, poorly defined and may vary in different parts of the protein. Moreover, dynamic effects may be significant. Therefore, we tend to trust the results obtained with atomistic simulations, like QTCP, which includes explicit dynamics and avoids the use of a protein dielectric constant.

Finally, we performed quantum refinement of the crystal structure with the four protonation states of homocitrate. We employed the raw data (the structure factors) of the starting crystal structure[48] and re-refined it with each of the four investigated protonation states of homocitrate. The four refinements were evaluated based on the maximum absolute RSZD score of the homocitrate ligand. The RSZD scores again showed that the 1Ha protonation state fits the crystallographic raw data best, slightly better than 2H state and much better than

the other two states. This is best seen in the electron maps, depicted in Figure 4.3b, which show that the O1 atom in the 2H structure is not in the right position, giving rise to significant difference densities.

a



b

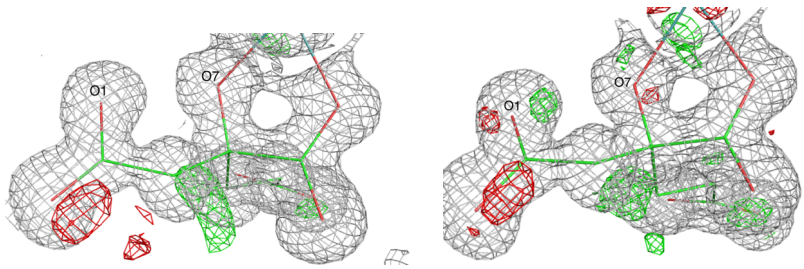


Figure 4.3: a) The four considered protonation states for homocitrate. b) Electron-density maps of the 1Ha (left) and 2H (right) protonation states of the homocitrate ligand in the quantum refinement of the nitrogenase 3U7Q crystal structure.[48] The $2mF_o - DF_c$ maps are contoured at 1.0 σ and the $mF_o - DF_c$ maps are contoured at +3.0 σ (green) and –3.0 σ (red).

4.2. Paper II

QM studies of nitrogenase are obstructed by the complicated electronic structure of the FeMo cluster. Experimentally, it is known to have three unpaired electrons in the resting E_0 state.[3] However, the individual Fe ions are in the high-spin Fe(II) or Fe(III) states, each with four or five unpaired electrons, respectively,[112] and the Mo ion is in a non-Hund Mo(III) state.[66] In QM calculations with DFT, the FeMo cluster is normally treated with the broken-symmetry (BS) approach,[122] applying separate wave functions for electrons of α and β spin.

In Section 3.1.5, previous studies of the BS states of the FeMo cluster were discussed, concluding that most studies have employed the BS7 state.[65,66,78,115,124,125] Unfortunately, it has been shown that the relative energies of the various BS states depend on the details of the QM calculations: Noodleman showed that the predicted ground state changed from BS2 to BS6 if the interaction energy with the surrounding protein and water was considered.[150] Moreover, different DFT functionals gave spin densities on the Fe ions that differed by over 0.6.[66]

To avoid the computationally demanding study of all 35 BS states for every putative intermediate in the reaction mechanism in nitrogenase, we developed a practical procedure to deal with the BS states in QM calculations of nitrogenase in Paper **II**. To achieve this goal, we studied systematically all 35 possible BS states of the FeMo cluster in nitrogenase with QM/MM methods for the resting state, as well as a one-electron reduced state and a singly protonated state with a proton on S2B atom. For each state, more than one spin state was considered (doublet and quartet for the resting and protonated states, singlet, triplet and quintet states for the reduced cluster). We examined how the relative energies of BS states are affected by

- the basis set (we tested two different basis sets of increasing size, def2-SV(P)[151] and def2-TZVPD)[152]
- the DFT functional (we employed two methods, TPSS[18] and B3LYP,[11,22,153])
- the geometry
- the surrounding protein and solvent

Our results indicated that

- The effect of the basis set is quite small (less than 11 kJ/mol) and the correlation (R^2) between the results obtained with the two basis sets is good 0.92–0.97. Therefore, initial evaluation of structures and spin states can be performed with a split-valence basis set.

- The DFT functional has a quite large effect on the relative energies of the various spin and BS states (up to 58 kJ/mol) and the correlation is rather poor, 0.57–0.72. Therefore, both a pure and a hybrid functional should be used to decide which BS state is most stable.
- Single-point calculations on one geometry give good correlations to results obtained with optimised structures, 0.92–0.98, but the results favour the BS state for which the geometry was obtained. Thus, a first scan of the energies of the various BS states can be performed without optimising the geometries.
- For all three states studied (resting, reduced and protonated states), BS7 was found to be most stable with both functionals and they also predicted that the most stable spin state was the quartet for the resting state and the quintet for the reduced state. The second most stable BS state varied depending on the applied functional.
- The 3–6 BS states of the same C_3-symmetry type (shown in Figure 3.6) had similar energies, within 14 kJ/mol. Thus, the protein has a rather small influence on the relative energies of the BS states related by the approximate three-fold symmetry of the FeMo cluster. This means that for initial investigations, it is enough to study only one example of each of the 10 symmetry-distinct BS states.

Based on these results, we suggested the following procedure for the study of the reaction mechanism of nitrogenase:

1. For each new protonation or oxidation state, all possible conformations should be studied by QM/MM geometry optimisation with DFT and medium-sized basis sets (e.g. TPSS-D3/def2-SV(P)), using the BS and spin state expected to be best (i.e. the one that was most stable for the previous state).
2. For the best structures, check the energies of the various BS states and other possible spin multiplicities. This can be done without geometry optimisation, with the same medium-sized basis set and only for the 10 symmetry distinct BS states, but both a pure and a hybrid functional should be used. If several BS states are degenerate within ~20 kJ/mol, geometry optimisation and calculations with larger basis sets are needed.
3. If a different BS or spin state is found to be most stable, the procedure needs to be repeated.
4. If the two DFT functionals give differing results regarding the most stable BS state, both states need to be examined and the energy difference indicate the uncertainty in the calculations.

5. Finally, an accurate energy should be calculated with a large basis set and it can be determined which of the 3 or 6 non-symmetric BS states is most stable.

The study gave the correct quartet ground state for the resting state of the FeMo cluster, providing some credence to the approach. For the reduced state, we find the quintet state most stable, in accordance with the results of Kästner and Blöchl.[114] That study also suggested that all intermediates are in the BS7 or BS6 states. This agrees with our results with the TPSS functional, showing that these two states are lowest for all three states studied. However, our results indicated that this conclusion may change if a hybrid functional is used instead. Therefore, you should always compare results obtained with several DFT methods; if they agree, the results can probably be trusted, whereas if they disagree, the difference reflects the uncertainty in the calculations. Our and others experience indicate that calculations with a pure and a hybrid functional or with two hybrid functionals with different amounts of Hartree–Fock exchange give a fast indication whether the results depend on the functional.[39,154]

4.3. Paper III

Spectroscopic, photophysical and kinetic experiments have suggested that the conversion of N_2 to NH_3 by nitrogenase involves a reductive elimination mechanism, i.e. that the N_2 binding is accompanied with the generation of a H_2 molecule at the E_4 state.[3,91,155] Consequently, a crucial question is the atomic structure of the E_4 state. Unfortunately, there is still no consensus regarding this structure. Based on ENDOR experiments and DFT calculations, Hoffman and coworkers have proposed that it contains two hydrides bridging between Fe2/6 and Fe3/7 and two protons binding to S2B and S3A (shown in Figure 3.4).[3,91,155] Predictions from computational studies have been strongly diverging,[89,91,105,116–118,156,157] as was discussed in Section 3.1.4. In particular, Siegbahn has argued that the experimentally suggested structure is not correct, because structures with protons on the centre carbide are much more stable.[117]

In an attempt to solve this controversy, we performed in paper III a systematic study to decide the most stable protonation state for the E_0–E_4 states of nitrogenase, based on the protein we set up in Paper I and the procedure to deal with BS states, designed in Paper II.

Addition of electrons is easy in the QM calculations, because they will automatically go to the proper position, during the wave function calculations. However, protons will normally stay on the atom where they are put at the start of the calculations, even if the structure is not the best. In the FeMo cluster, a proton can be added to at least 21 atoms (Mo, seven Fe ions, nine S ions, the carbide ion, as well as the cysteine, histidine and homocitrate ligands) and at each site, the proton may point to several different directions, leading to a total of more than 50 possible positions. Moreover, for each structure, there are 35 possible BS states and 2–4 spin states needed to be tested. This means that if we want to predict the E_4 state without any compromise, we need to test of the order of 10^9 different structures ($50^4 \times 35 \times 4$), which is out of the reach of current computational recourses. Therefore, all studies must follow some sort of heuristic approach to find the best protonation states. We decided to systematically add protons to all ~50 potential protonation sites, starting from the most stable structure of the state with one proton and electron less.

As usual, we employed the QM/MM approach, with two DFT methods, a pure functional TPSS and a hybrid functional B3LYP, and two different basis sets of increasing size, def2-SV(P) and def2-TZVPD. For each E_n state, we tested ~50 possible protonation states. For the best candidates, we checked the energies of all 35 BS states and tested also several spin states.

With such an approach, it is necessary to first settle the protonation state of E_0. To that end, we employed quantum refinement, comparing the various protonated

structures in terms of the RSZD score and the electron-density difference maps. Our calculations clearly showed that E_0 is unprotonated. The difference-density maps of the quantum-refined unprotonated and the best protonated structure from the DFT investigation show clear positive and negative difference density peaks next to the protonated S2B atom, as can be seen in Figure 4.4 (marked with red arrows). This was further supported by the RSZD score of the S2B atom, going from –0.3 in the unprotonated structure to –2.6 in the protonated structure.

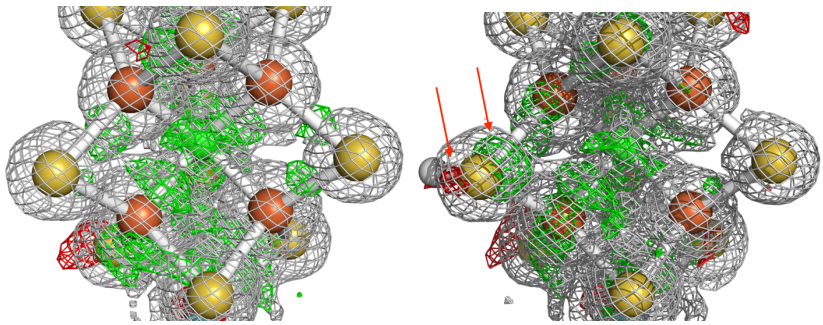


Figure 4.4: Electron-density maps of the resting E_0 state without (left) or with a proton (right) on the S2B atom (middle left side) in the quantum refinement of the nitrogenase 3U7Q structure.[48] The $2mF_o - DF_c$ maps are contoured at 1.0 σ and the $mF_o - DF_c$ maps are contoured at +3.0 σ (green) and –3.0 σ (red). The red arrows point out the extra difference density for the protonated structure.

For the E_1 state, we find that the most favourable protonation site is on S2B, pointing towards S3A, as shown in Figure 4.5. This agrees with most previous theoretical studies.[3,66,98,103–114] We also decided the most stable BS state and confirmed that the quintet was the most stable spin state.

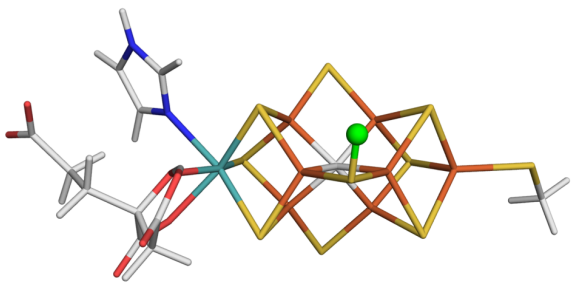
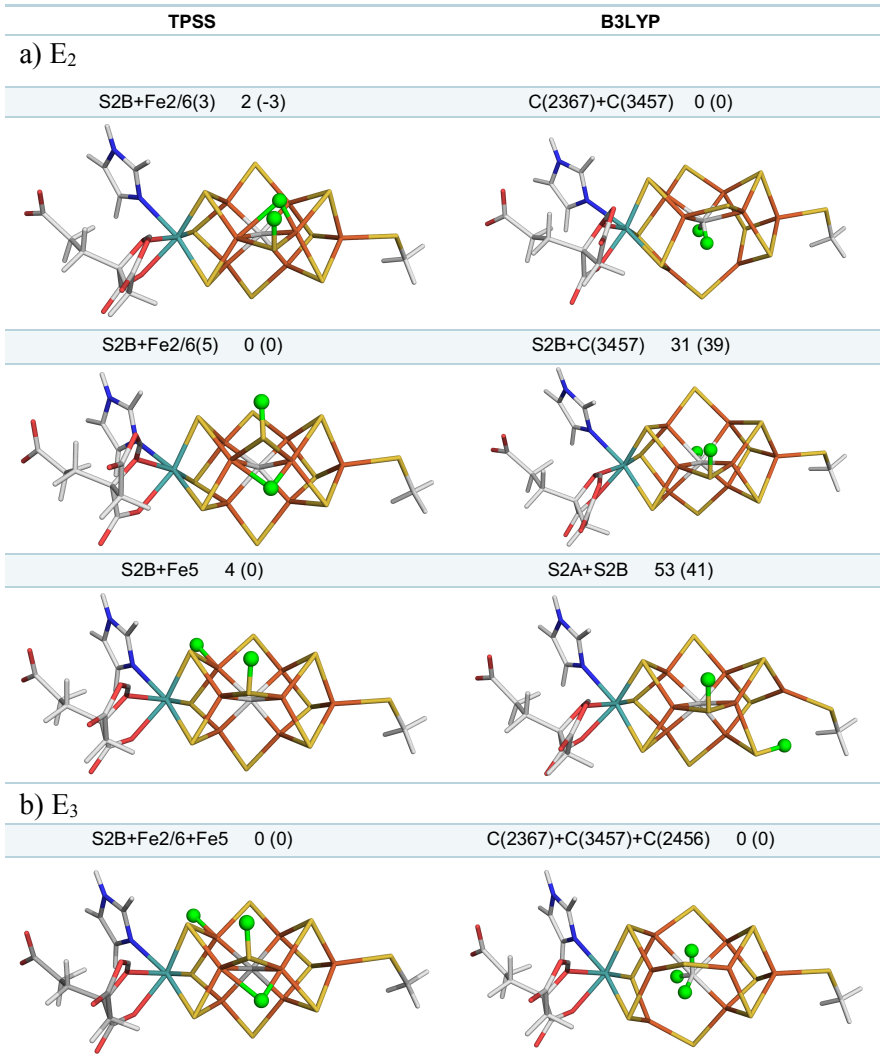


Figure 4.5: The best E1 structure, protonated on the S2B atom, obtained at the TPSS-D3/def2-SV(P) level of theory.

For the E_2–E_4 states, we found new preferred protonation states that had not been suggested before. Moreover, the predictions strongly depended on which DFT method was used to calculate the relative energies. The TPSS functional favoured structures with hydride ions on the Fe ions, i.e. the most stable structures involved

hydride ions bridging the Fe2 and Fe6 ions and terminally binding to Fe4, Fe5 or Fe6. In each E_n state, several structures were close in energy, indicating that the hydride ions may move quite freely within the cluster. On the other hand, B3LYP disfavoured metal-binding hydride ions and instead strongly preferred protonation of the central carbide ion. This is in agreement with the suggestions by Siegbahn,[105,117] although the carbide ion stayed in the centre of the (quite distorted) FeMo cluster in our QM/MM calculations. The most stable structures are displayed in Figure 4.6. In agreement with our previous study, the energies are insensitive to the size of the basis set (changing by less than 16 kJ/mol going from def2-SV(P) to def2-TZVPD).



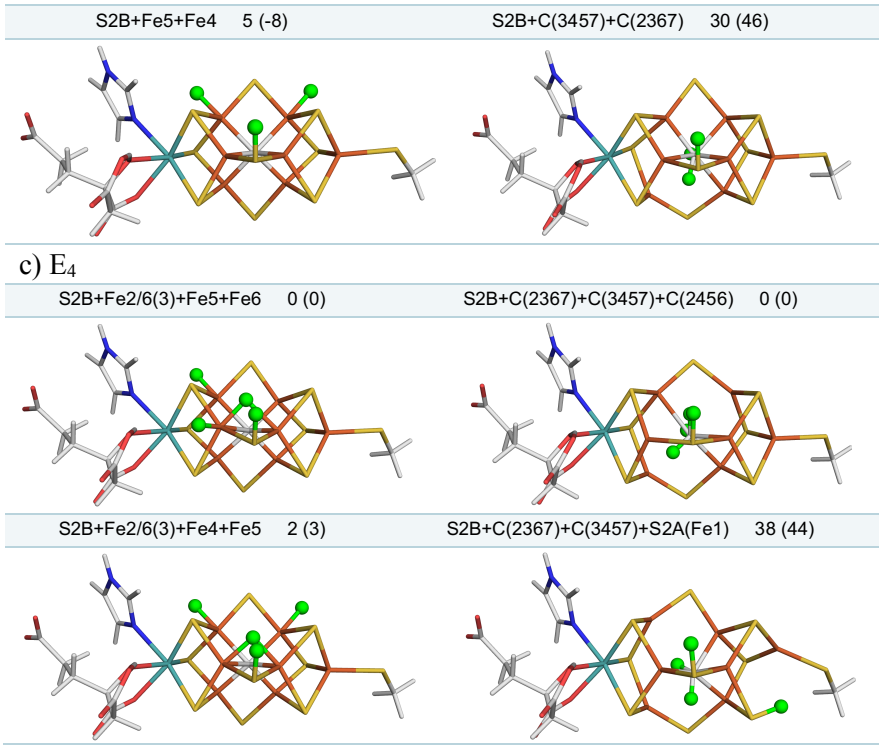


Figure 4.6: The most stable a) E_2, b) E_3 and c) E_4 structures obtained at the TPSS-D3/def2-SV(P) (left column) and the B3LYP-D3/def2-SV(P) levels of theory (right column). For each structure, the relative energies (kJ/mol) are shown to the right (values in parenthesis are from single-point calculations with the TZVP(D) basis set).

We also studied the best E_4 structures suggested by Hoffman Adamo, Dance and McKee and coworkers.[106,107,155,158,159] The energies are compared to our best structures in Table 4.1. It can be seen that our structures are always significantly better (by at least 52 kJ/mol). It can also be seen that the relative energies obtained with TPSS and B3LYP are very different, e.g. differing by 308–389 kJ/mol for the best structures obtained with TPSS or B3LYP.

Table 4.1. Comparison of our best E_4 structures with those suggested by Hoffman Adamo, Dance and McKee and coworkers.[106,107,155,158,159]

Structure		**def2–SV(P) (kJ/mol)**	
		TPSS	**B3LYP**
Hoffman	S2B+S5A+Fe26+Fe37	52	264
	S2B+S5A+Fe2/6+Fe2	84	365
	S2B+S3A+Fe2/6+ Fe3/7	131	412
Dance	S2B+S3B+Fe2+Fe6	72	426
Adamo	S2B+S2A+S5A+C(3457)	288	146
Paper III	S2B+Fe2/6+Fe5+Fe6	0	389
Paper III	S2B+C(2367)+C(3457)+C(2456)	308	0

4.4. Paper IV

Many computational studies of nitrogenase with DFT methods have been published, but they do not show any consensus – they do not even propose the same protonation states for the central intermediate E_4.[4] In Paper IV we show that the prime reason for these diverging predictions is that different DFT methods were applied in the various studies. As already was observed in Paper III, the hybrid functional B3LYP prefers to add protons to the central carbide ion, whereas the pure GGA functional TPSS prefers structures involving hydride ions bound to the Fe ions.

The purpose of Paper IV was to describe and understand the large variation of the results obtained with the various DFT methods and try to decide which methods describe nitrogenase best. To this end, we employed ten models of the FeMo cluster in the E_4, E_2 or E_0 state, shown in Figure 4.7. We investigated their dependence on the DFT method by comparing the energies and the geometries obtained from QM/MM optimisation calculations with 13 different DFT methods: the (meta) GGA functionals TPSS, PEB, BP86, BLYP, M06-L and B97D and the hybrid functionals TPSSH, B3LYP, PBE0, M06, BHLYP, M06-2X and M06-HF.[11–23]

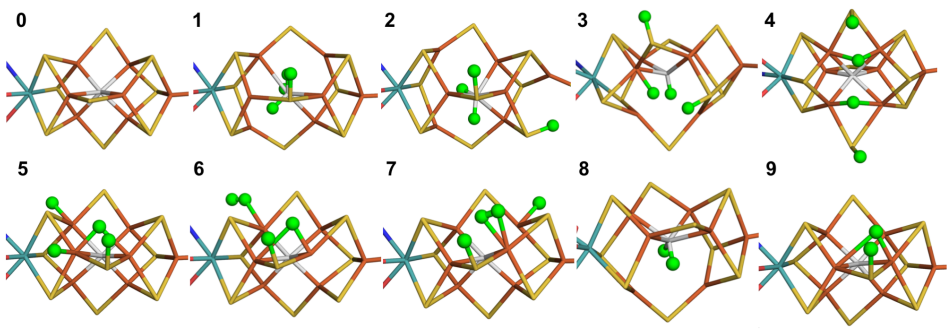


Figure 4.7: Structures of the ten studied nitrogenase models, **0–9**. Added hydrogen atoms are shown in green balls.

To understand the energy differences, we first compared the relative stability of the best protonation states of the E_4 state obtained with either the B3LYP (**1**) or TPSS (**5**) functionals in Paper III, ΔE_{15}. We performed QM/MM optimisations of **1** and **5** with each of the 13 DFT methods. Our calculations showed that structure **5** was 8–242 kJ/mol more stable than structure **1** with (meta) GGA method whereas structure **1** was 39–855 kJ/mol more stable than structure **5** when the hybrid functionals were applied. Thus, the relative stability predicted by different methods differed up to 1097 kJ/mol.

Since structure **1** contains a triply protonated central carbide ion but no Fe-bound hydride ions, whereas structure **5** contains no protons on the central carbide but three hydride ions, we introduced structures **2, 3** and **4,** which contain 2, 1 and 0 protons on the central carbide and 0, 1 and 2 metal-bound hydride ions, and performed the same calculations on them, to further understand the stability dependency on the various DFT methods. We found that the discrepancy between the relative stability predictions by the 13 DFT methods was moderate for ΔE_{12} and ΔE_{45} but large for ΔE_{23} and ΔE_{34}. This indicates that the large variation in ΔE_{15}, ΔE_{23} and ΔE_{34} can be explained by at least two aspects. The first is the difference in the formal oxidation state of the Fe ions in the cluster. A hydride ion contains two more electrons than a proton, so the formal oxidation state of the Fe ion changes when a hydride ion binds to it. Thus, the formal oxidation state of the Fe ions in structure **5** with three hydride ions has increased by six compared to structure **1** with no Fe-bound hydride ions. The second, is the number of Fe–S and Fe–C bonds. The addition of protons to the central carbide ion leads cleavage of Fe–C bonds and to a distortion of the cluster. The change in the number of Fe–C bonds in the **2** → **3**, **3** → **4** and **1**→ **5** transitions is 2, 1 and 4 . The largest effects of the functionals are seen when both the oxidation state and the number of Fe–C bonds change.

To gain additional information about how the energies vary with the DFT functionals, we also studied the binding and dissociation of H_2 from the E_4 (structure **5**/**1**) and E_2 states (structure **9**/**8**, i.e. the best E_2 structures obtained in Paper III with B3LYP and TPSS, respectively). We showed that all methods except B3LYP and PEB0 gave strongly favourable H_2 dissociation energies from E_4 and E_2 state, which probably overestimates the risk that H_2 dissociates before N_2 binds to the cluster.

We also studied which of the DFT functionals produce the best geometries by comparing the QM/MM optimised structures of the E_0 state to the crystal structure of nitrogenase (3U7Q) in the resting state.[48] Based on the RMSD, as well as the mean absolute deviation (MAD) and maximum deviation of the metal–metal and the metal–ligand distances, we found that the GGA functionals (except M06-L) and TPSSh gave the best structures. Thus, we showed that it is currently unclear which DFT method gives the best results for nitrogenase. Non-hybrid DFT functionals and TPSSh give the most accurate structures of the resting active site, whereas B3LYP and PBE0 give the best H_2 dissociation energies. However, none of the calculations indicated that a E_4 structure with two bridging hydride ions is lowest in energy, as suggested by ENDOR spectroscopy.[3,88,89]

4.5. Paper V

As mentioned before, it is believed that the FeMo cluster needs to accept four protons and electrons, to reach the E_4 state, before it can bind N_2 and convert it to NH_3.[3] However, there is still no consensus on the atomic structure of the E_4 state. Based on ENDOR experiments, Hoffman and coworkers have suggested that two of the added protons bind to the sulfide ions whereas the other two bridge between two pairs of Fe ions as hydride ions.[3,88] With the help from DFT calculations, they have suggested the structure in Figure 3.4,[89] i.e. that all four added protons bind to the same face of the cluster, with protons on the S2B and S5A sulfides, and hydride ions bridging the Fe2/6 and Fe3/7 pairs. However, in Paper III, we performed a thorough study of possible protonation states and showed that this is not the structure of E_4 with the lowest energy. Instead, our results suggested that the E_4 structure contains either three hydrides ions binding to Fe ions (as predicted by TPSS) or a triply protonated central carbide ion (as predicted by B3LYP), cf. Table 4.1. Moreover, in Paper IV, we performed a more thorough study with 13 different DFT methods. Still, none of these studies pointed out any structure with two bridging hydride ions, as the most stable E_4 structure. This is a problematic discrepancy between experiments and computational studies.

However, as mentioned above, there are at least 50^4 = 6 million possible protonation states of the E_4 structure and we tested only ~150 of these (albeit with a heuristic and systematic approach, starting from the best E_3 states). Therefore, it is possible that we missed some important structures. Consequently, we made a more focused study in Paper V. We assumed that the ENDOR results are correct and concentrated on structures with two bridging hydride ions and two protonated sulfides.

There is still a very large number of such structures and to reduce it to a manageable number, we started by deciding the best protonation sites for the sulfide ions. Since essentially all previous studies agree that the first proton (in E_1) should bind to S2B,[4] we always kept this ion protonated. Then, we compared the relative energies of the 16 possible structures obtained by varying the position of the second protonated sulfide ion, all with S2B protonated and with two hydride ions bridging between Fe2/6 and Fe3/7, as in the Hoffman structure. Since the experiments suggest bridging hydride ions and such states are disfavoured by hybrid functionals, we employed the TPSS-D3/def2-SV(P) method. The calculations indicated that the best position of the second proton is on S5A, pointing towards S3A. This is similar to the Hoffman structure, but the proton on S5A points in the opposite direction (not towards S2B as in Hoffman structure).

Next, we started a systematic search for the best E_4 state with two bridging hydride ions, retaining the two protons on S2B and S5A. There are 12 possible binding positions for one bridging hydride, so we tested all 66 ($12 \times 11 / 2$) possible structures with two bridging hydride ions. Our results indicated that the most stable E_4 structures have two hydride ions bridging between Fe2 and Fe6 and between Fe3 and Fe7, as suggested by Hoffman and coworkers. This indicates that calculations based on DFT methods actually start to converge toward the experimental observations. However, we find that the hydride ion between Fe2 and Fe6 is more favourable on the other side of S2B, i.e. pointing towards S3A, rather than S5A.

We also showed that the positions of the two protons may further affect the relative energy. We performed some additional variations among the best structures, by changing the conformation of the protons on S2B and S5A, or by moving the proton on S5A to S3B. By such variations we identified the most stable structure, shown in Figure 4.8a. However, the relative energies were sensitive to the details of the calculations, and if the surrounding protein was allowed to relax, the structure with the opposite orientation of the two protons on the sulfide ions became more stable, shown in Figure 4.8b.

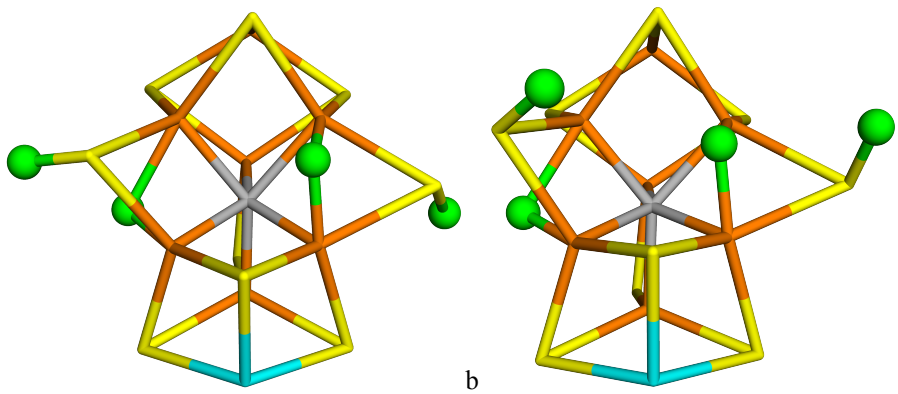


Figure 4.8: The two most stable structures a) S2B(3)−S5A(3)−Fe2/6(3)−Fe3/7(2) and b) S2B(5)−S5A(2)−Fe2/6(3)−Fe3/7(2).

To investigate how the relative energies depend on different DFT methods, we employed seven additional functionals, viz., PBE, M06-L, B97D, TPSSh, B3LYP, PBE0 and M06. With these functionals, we compared the energies of the best TPSS structure in Paper IV (called 3H), with the best eight E_4 structures in Paper V. Interestingly, all eight functionals showed that the 3H structure was 1–93 kJ/mol less stable than the two most favourable E_4 structures with two bridging hydride ions. An important reason for the preference of the latter structures was that they attained a new BS state, involving only two Fe ions with

minority spin rather than three as was assumed in Paper I–IV and in all previous DFT studies of nitrogenase.[4] This new BS state reduced the energy of the best E_4 structures by ~19 kJ/mol. Consequently, this study indicates that the pure functionals actually do suggest a E_4 structure in accordance with experiments, whereas the hybrid functionals still prefer structures with the central carbide ion triply protonated.

Moreover, the Hoffman structure was not the most stable state for any of the seven functionals. However, with other conformations of the two protons on the sulfide ions, we could obtain Hoffman-like structures (i.e. with hydride ions on Fe3/7 and on Fe2/6 directed towards S5A) that are within 6–22 kJ/mol of our best structure in Figure 4.8a. This may be within the uncertainty of the present study, considering that the Hoffman-type structures reproduce details of the ENDOR measurements[89] better than the structure in Figure 4.8a.

4.6. Paper VI

In Papers I–V, we thoroughly investigated the protonation states of the residues around the cluster and the most stable structures in E_0–E_4 states. We pointed out the importance of finding a heuristic but systematic approach to make the calculations possible and reproducible, and confirmed that the approach we suggested gives results close to experiments. Therefore, in this study, our research proceeded to the next stage of the mechanism, N_2 binding.

It is believed that N_2 binds to the cluster after the addition of four protons and electrons to the cluster (E_4 state). Moreover, it is also assumed that during the binding of N_2 to the cluster, H_2 dissociates [3] and that N_2 immediately takes up the remaining two protons on the cluster to form N_2H_2.[3,91] Thereby, the FeMo cluster is brought to the same redox level as in the E_0 state, with no extra protons. These two reactions strongly simplify our study. The geometry, spin and BS state of the resting state are known from previous experimental and computational studies.[3,4] With no protons bound to the cluster, the millions of possible protonation states of E_4 is no longer a case. All this enabled us to study systematically all possible N_2H_2-bound structures.

Thus, we tested end-on binding modes of NNH_2, side-on binding modes of cis-HNNH, as well as end-on binding modes of HNNH, in both the cis and trans conformations (cis-HNNH has the two hydrogen atoms on the same side of the N–N bond). We obtained 21, 30, 6 and 7 different structures for the four binding modes, differing in how many and which metal ions N_2H_2 binds to. All structures were studied for the quartet state, as is observed for E_4[3] and they were optimised with QM/MM approach, using both TPSS-D3/def2-SV(P) and the B3LYP-D3/def2-SV(P) methods. In addition, single-point energies were also calculated with the large def2-TZVPD basis set. For some structures, we also allowed the surrounding protein free to relax.

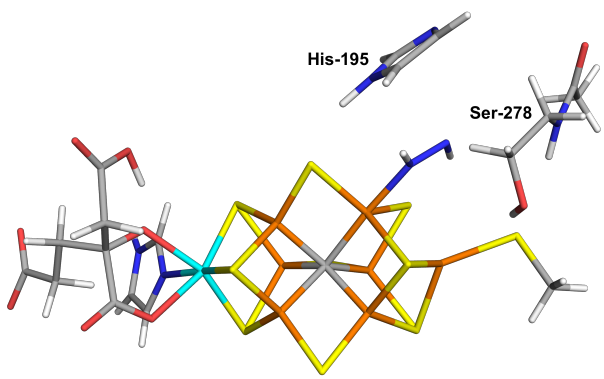


Figure 4.9: The most stable binding mode found in paper VI, with trans-HNNH binding to Fe2.

Our results indicated that the binding of N_2H_2 is determined more by interactions and steric clashes with the surrounding protein than by the intrinsic preferences of the ligand and the FeMo cluster. The best binding mode with both the TPSS and B3LYP DFT methods had trans-HNNH terminally bound to Fe2. It is stabilised by stacking of the substrate between His-195 and Ser-278, as can be seen in Figure 4.9. However, several other structures came rather close in energy (within 3–40 kJ/mol) at least in some calculations. The corresponding structure with cis-HNNH terminally bound to Fe2 was second best with B3LYP. A structure with $HNNH_2$ terminally bound to Fe6 (the third proton was taken from homocitrate) was second most stable with TPSS and also if the surrounding protein was allowed to relax. This structure was stabilised by hydrogen bonds with the alcoholic proton and the acetate group of homocitrate. With the TPSS functional, a structure with cis-HNNH side-on binding to the Fe3–Fe4–Fe5–Fe7 face of the cluster was also rather low in energy, but all side-on structures were strongly disfavoured by the B3LYP method.

Table 4.2 shows the relative energies of the six best binding modes predicted by five methods. It is satisfying that our six best structures all involve N_2H_2 binding to Fe2 and Fe6, because experimental observations have suggested the Fe2–Fe3–Fe6–Fe7 face as the reactive side of the cluster.[3,160,161]. This shows that our QM/MM approach works properly and is accurate enough to find the most reactive sites without restricting the search by experimental information. Consequently, our structures are the most likely candidates for the N_2-bound structure of nitrogenase.

Table 4.2. Relative energies of the six most favorable binding modes of N_2H_2 binding to the FeMo cluster. The five energies are TP – TPSS-D3/def2-SV(P) optimised geometries with surroundings fixed, TZ – single-point TPSS-D3/def2-TZVPD on the TP structures, Free – TPSS-D3/def2-SV(P) optimised geometries with the surroundings relaxed, B3 – B3LYP-D3/def2-SV(P) optimised geometries with surroundings fixed and B3Free – B3LYP-D3/def2-SV(P) optimised geometries with the surroundings relaxed.

	TP	TZ	Free	B3	B3Free
End-on					
Fe2(trans)	0	0	0	0	0
Fe2(cis)	17	12	86	18	46
Fe6(HNNH2)	9	3	23	29	50
Fe6(HCA)	38	28	52	68	51
Fe6	56	44	91	90	70
Side-on					
Fe3/7Fe4/5	40	31	85	186	192

4.7. Paper VII

Most studies have assumed that the FeMo cluster remains intact, i.e. in the $MoFe_7S_9C$ state, throughout the reaction. However, two recent crystal structures of Mo and V-nitrogenase have indicated that the S2B atom can reversibly dissociate from the cluster, leaving a vacant binding site for the substrate.[70,99] This is an interesting observation, because it would immediately suggest the binding site of N_2, thereby avoiding a lengthy investigation of the N_2 binding mode, like the one performed in Paper VI. Currently, the research field is divided into two camps regarding the significance of this finding: Some scientists, believe that the crystal structures have revealed the actual reaction mechanism of nitrogenase, involving dissociation and re-binding of S2B.[136] For example, already in 2015, Nørskov et al. suggested a mechanism for Mo-nitrogenase in which a doubly protonated S2B atom was released, based on a minimal model of the cluster and a doubly protonated resting state[98] (which is incorrect according to our calculations in Paper IV). In their mechanism, only one of the N atom bridges between Fe2 and Fe6, forming the first NH_3 product by sequentially protonating the distal N atom. This is one of two possible reaction paths discussed for nitrogenase[3,100] (Section 3.1.3), shown in Figure 3.5.

Other scientists suppose that the crystal structures show inhibited states that are not along the normal reaction mechanism.[89,162] In fact, recent DFT calculations have indicated that dissociation of S2B would involve too high barriers (>80 kJ/mol) to be kinetically feasible,[163] although another study indicated that the strength of the Fe–S bonds depends on the DFT method.[162] Interestingly, both these studies instead suggested that S2B may dissociate from only one of the two Fe ions, still forming a N_2 binding site between the two ions. With this lack of consensus, we decided to keep both possibilities open and to investigate whether they may lead to credible reaction mechanisms with favourable energetics.

In Paper VII, we performed the first part of such a study, by studying whether a reaction mechanism for the Mo-nitrogenase with a dissociated S2B group is thermodynamically feasible and whether the NH_3 formation follows an alternating or distal mechanism. Our calculations are appreciably more accurate than those of Nørskov and coworkers:[98] All structures were investigated with QM/MM approach, optimised with TPSS-D3/def2-SV(P) and B3LYP-D3/def2-SV(P) methods.

Like in Paper VI, we assumed that the N_2 substrate is protonated to N_2H_2 immediately after its binding to the cluster. The S2B group was removed from the QM system and Gln-191 was rotated as indicated in the crystal structure of V-nitrogenase.[99] For each intermediate E_4–E_8, we tested all structures that may be involved in the alternating or distal mechanism, considering also different

protonation states of His-195, which may form hydrogen bonds with the substrate and the intermediates.

Based on our results, we proposed a mechanism that is mainly alternating, as displayed in Figure 4.10. For the E_5–E_7 states, the protons are added alternatively to the two N atoms. However, for the E_4 state, we actually find that the NNH_2 structure, rather than HNNH is most stable, albeit the former intermediate traditionally is connected to the distal reaction mechanism. For the E_5 state, we find that a H_2NNH structure is 51 kJ/mol more favourable than the NNH_3 intermediate. This represents the crossover to the alternating mechanism. For the E_6 state, a rather symmetric H_2NNH_2 (hydrazine) structure is most stable. It is protonated in the E_7 state and this leads automatically to the cleavage of the N–N bond. However, dissociation of the formed NH_3 product requires one extra protonation and reduction. The second NH_3 is formed at the E_8 stage. Interestingly, its dissociation is accompanied with the rebinding of S2B to the cluster and Gln-191 rotates back to its original position, reforming the resting E_0 state of the cluster.

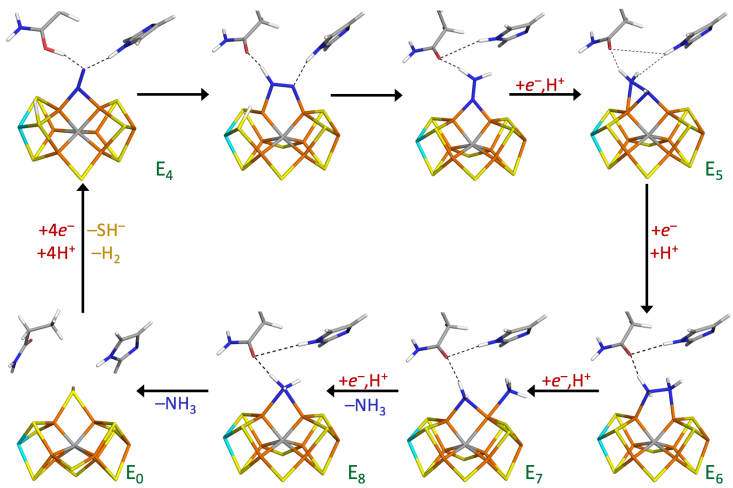


Figure 4.10: Our putative reaction mechanism of Mo-nitrogenase with a dissociated S2B, based on the QM/MM calculations in paper VII.

Thus, our calculations show that formation of NH_3 by Mo-nitrogenase with a dissociated S2B ligand is thermodynamically feasible. It follows a mainly alternating mechanism with NNH_2, $HNNH_2$ and H_2NNH_2 as the energetically most stable intermediates for the E_4, E_5 and E_6 states. Still, this does of course not prove that this is the actual mechanism followed by Mo-nitrogenase. This would require similar studies for the corresponding reaction without dissociation of S2B, as well as detailed studies of the mechanism for the dissociation of S2B and the binding of N_2.

4.8. Paper VIII

Throughout Papers II–VII, our studies were concentrated on the catalytic cluster of nitrogenase. However, the full reaction mechanism of the enzyme also involves the P-cluster, which transfers electrons to the catalytic cluster. As mentioned in Section 3.2, the P-cluster is a $Fe_8S_7Cys_6$ complex. Quite unexpectedly, the resting state is the fully reduced ($Fe(II)_8$; P^N) cluster and states oxidised by 1–3 electrons (P^{1+}, P^{2+} and P^{3+}) have been studied. At the start of our study, crystallographic structures were available only for the P^N and P^{2+} states, showing that in the latter, two residues are deprotonated (the side-chain oxygen atom of Ser-D188 and the backbone N atom of Cys-C88) and forms bonds to two Fe ions (Fe6 and Fe5, respectively). However, the structure of the P^{1+}state was not known and experiments indicated that the transition between the P^N and P^{1+} states does not involve any proton transfer.[133,134] Yet, while working on Paper VIII, a crystal structure of a putative P^{1+} state was published.[47]

The aim of Paper VIII was to make computational studies of the P-cluster possible. For that, two problems had to be solved. First, the exact protonation state of all oxidation states needed to be settled. Second, the proper electronic structure, including the spin states and the BS states had to be found. For the first question, we tested four different protonation states: both Ser-D188 and Cys-C88 protonated, both residues deprotonated or one of them protonated and the other deprotonated. For the P^N and P^{2+} states, a high-resolution structure was available (1.0 Å resolution),[48] and by comparing our QM/MM structures to this, we could unambiguously show that P^N has both Ser-D188 and Cys-C88 protonated and P^{2+} has both Ser-D188 and Cys-C88 deprotonated, even if the accuracy of the QM/MM structures is somewhat worse than for the FeMo cluster, because the crystal structure is a mixture of the P^N (20%) and P^{2+} (80%) states, with differing coordinates only for two atoms, Fe5 and Fe6. The crystal structure of the putative P^{1+} state[47] is of a much lower quality (2.1 Å). We showed that it involves a mixture of the P^{1+} and P^{2+} states, and that several of the reported Fe–Fe and Fe–S distances in the cluster are highly questionable. Still, we could confirm the suggestion of the crystallographers that the P^{1+} state involves a deprotonated Ser-D188 and a protonated Cys-C88, shown in Figure 3.7c.

Once, the protonation states were settled, we decided the best spin and BS states of the P-cluster in the P^N, P^{1+}, P^{2+} and P^{3+} oxidation states. Based on experimentally reported spin states[137,139] and an interpretation of these,[137] we systematically went through all possible BS states (16–48) with one, two or three different coupling schemes. For the P^N state, both TPSS and B3LYP agreed that the best BS state has minority spin on the Fe1, Fe2, Fe4 and Fe7 ions and this is also the state that reproduced the crystal structure best (Figure 4.11). Quite satisfactorily, this state is supported by a recent density-matrix renormalisation-

group complete active-space study,[138] which provides a theoretically more correct description of the electronic structure, but not more accurate relative energies of the various BS states (dynamic correlation is not accounted for). For the other three oxidation states, the two DFT methods gave diverging predictions regarding the BS and spin states. However, we tend to suggest that P^{1+} is best modelled by the same BS state as P^{N}, whereas the lowest BS state for both P^{2+} and P^{3+} have minority spin on Fe3, Fe5 and Fe8.

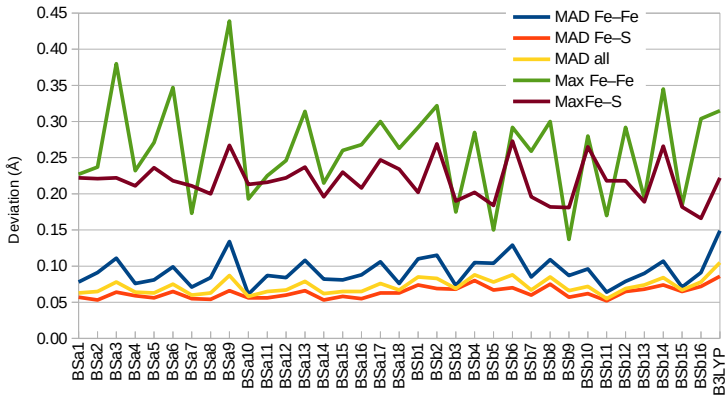


Figure 4.11: Comparison between the 3U7Q crystal structure[48] and the various BS states for the P-cluster in the P^{N} state. The figure show the mean absolute deviations (MAD) for the Fe–S, the Fe–Fe distances and both distances together (all), as well as the maximum deviation for the Fe–S and Fe–Fe distances. All results are obtained at the TPSS-D3/def2-SV(P) level, except the last entry in a) which shows the best BS (BSb11) state obtained with the B3LYP functional. The best state is BSb11, which has minority spin on the Fe1, Fe2, Fe4 and Fe7 ions.

4.9. Paper IX

X-ray crystallography is the main source of atomistic information for the structure of proteins. Owing to the limited resolution of the crystallographic data for proteins, the structures are obtained as a compromise between the X-ray scattering data and a set of empirical restraints, which ensure chemically reasonable bond lengths and angles (cf. Eqn. 2.40 in Section 2.4.1). However, for non-standard residues such as substrates, inhibitors and metal sites, empirical restraints are not always available or accurate. Such heterocompounds are typically found in the active site, which are the mechanistically most interesting part of the proteins. This can be solved by our quantum-refinement approach,[58,59] in which the empirical restraints for a site of particular interest are replaced by quantum-mechanical (QM) calculations (introduced in Section 2.4.2).

However, a limitation of standard quantum refinement has been that it does not allow disorder in the QM system. Unfortunately, such disorder is quite common in metal-containing protein active sites, e.g. owing to photoreduction during data collection.[164–166]. In Paper VIII, we observed such a problem for the crystal structure of the putative P^{1+} state of the P-cluster in nitrogenase: We showed that the crystal structure is most likely a mixture of the P^{1+} and P^{2+} states. Therefore, a quantum-refined structure of this protein will be a weighted average over the two states and therefore not accurate for any of the states.

To solve this problem, we extended in Paper IX our previous quantum-refinement approach to allow disorder in the QM region, by employing separate QM calculations on the two conformations. This was obtained by a pseudo-energy function of the form

$$E_{\mathrm{cqx}} = w_{\mathrm{A}}E_{\mathrm{Xray}} + w_{\mathrm{MM}}E_{\mathrm{MM}} + n_{\mathrm{occ1}}(E_{\mathrm{QM11}} - w_{\mathrm{MM}}E_{\mathrm{MM11}}) - n_{\mathrm{occ2}}(E_{\mathrm{QM12}} - w_{\mathrm{MM}}E_{\mathrm{MM12}}) \qquad (4.1)$$

Here, E_{Xray} and E_{MM} are the same as in Eqn. 2.40, but they now involve alternative conformations of atoms in the QM system, treated with standard methods in the crystallography software (e.g. to avoid interactions between the two conformations in the MM term). E_{QM11} and E_{MM11} are the QM and MM energies of the first conformation of the QM system (called system 11), which has the occupancy n_{occ1}. Likewise, E_{QM12} and E_{MM12} are the QM and MM energies of the second conformation of the QM system (called system 12), which has the occupancy n_{occ2}. w_{A} is the same weight factor as in Eqns. 2.40 and 2.41, determining the relative importance of the experimental and QM or MM data. w_{MM} is a weighting function needed because the empirical potential of crystallographic refinement software is based on statistics, rather than energies (as is the QM term). It is the inverse of the w_{QM} term in Eqn. 2.41. As for standard

quantum refinement, the method was implemented as an interface between the QM software Turbomole[61] and the crystallography software CNS. [62,63]

To illustrate the performance of this novel approach, called ComQumX-2QM, we applied it to the P-cluster in nitrogenase, based on two different crystal structures. The first is the 3U7Q structure,[48] which was obtained at a high resolution (1.0 Å) and contains two conformations of the P-cluster, interpreted as a mixture of the fully reduced state P^N (20% occupancy) and the two-election oxidised state P^{2+} (80% occupancy). In practice, however, the coordinates differ for only two atoms, Fe5 and Fe6. The second crystal structure is 6CDK,[47] which was obtained at a moderate resolution (2.1 Å) and which originally was suggested to contain only the P^{1+} state, although our studies in Paper VIII indicated that it is a mixture of the P^{1+} and P^{2+} states.

We started by performing the ComQum-2QM quantum refinement on the 3U7Q crystal structure, with a default weight factor of $w_A = 0.077$ and the original occupancies (20/80%). The quantum-refined structure reproduced the crystal structure well, illustrated by small mean absolute deviations (MAD) to the original crystal structure of 0.01–0.02 Å and 0.04–0.06 Å for Fe–Fe and Fe–S distances in the P^N and P^{2+} states, respectively. Next, we showed that we can vary the weights of X-ray and QM/MM restraints to get an optimal structure and that the *R* factors, RSZD scores, strain energies and electron-density difference maps could give a good validation of the weights. With the best value of w_A (0.1), we performed four occupancy refinements of the P-cluster and the coordinating residues. Taking the *R* factors, RSZD scores, strain energies and difference electron-density maps into account, we proposed a best quantum-refined structure composed of 15% P^N state and 85% P^{2+} state

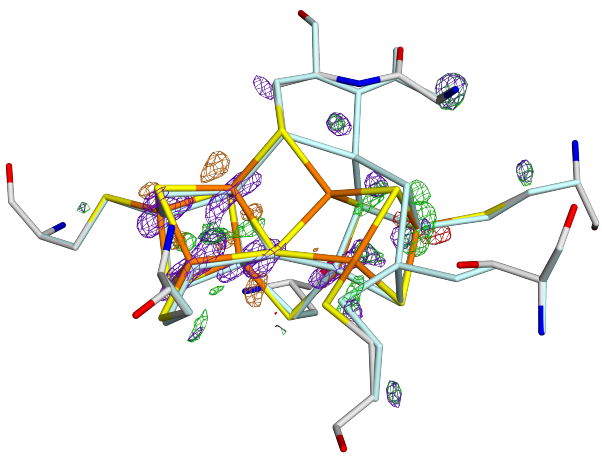


Figure 4.12: The $mF_o - DF_c$ electron-density difference maps at a ±4 σ level of the quantum-refined structures with w_A = 0.1 and occupancies 15/85% (green positive and red negative) compared to the original crystal structure of 3U7Q (purple and orange). The P^N state is shown in atomic colours, whereas the P^{2+} state is show in pale cyan.

The quantum-refined structures indicated that many atoms, besides Fe5 and Fe6, move significantly between the two oxidation states, by up to 0.70 Å. Yet, such differences are too small to be discerned by crystallography. The best quantum-refined structures show an improved description of the P-cluster with RSZD scores of 9, compared to 22 in the original crystal structure. The improvement is also reflected by the electron-density difference maps, as can be seen in Figure 4.12: The difference-density volumes for the original crystal structure (displayed in purple and orange in the figure) are larger than that for the quantum-refined structure (green and red). Thus, the quantum-refined structure describes the P-cluster much better than that in the original 3U7Q crystal structure.

For the 6CDK crystal, we used a similar procedure. The best quantum-refined structure, calculated with $w_A = 1$ indicated that it is a mixture of 50% P^{1+} and 50% P^{2+}, rather than only the P^{1+} state as originally suggested.[47] Unfortunately, the improvement for the electron-density difference maps is less conclusive for this structure: a big positive difference (purple) blob in the original structure has disappeared in our best quantum refined structure, but a positive difference (green) around Cys-88 and negative difference (red) around Fe6 have appeared, as shown in Figure 4.13. This is because of the lower resolution of this structure.

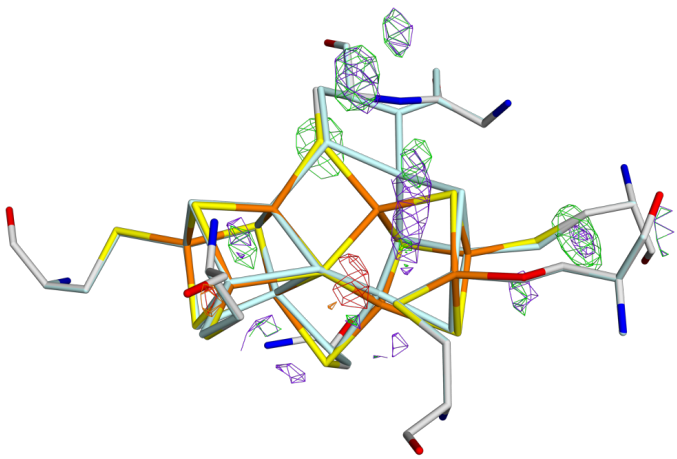


Figure 4.13: The $mF_o - DF_c$ electron-density difference maps of the best quantum-refined structure of 6CDK with $w_A = 1$ (positive density in green and negative in red) as well as the original crystal structure (purple and orange). All maps are contoured at the ±2.5 σ level. The P^{1+} state is shown with atomic colors and P^{2+} state in pale cyan.

We have also considered strain energies, which are the difference in QM energy of the QM system in the quantum-refined structure compared to that in the structure obtained without any crystallographic data (i.e. the pure QM/MM structure using the CNS MM force field). They reflect how much the current model is biased from the QM/MM structure, i.e. how much the structure is distorted by the crystallographic data (how well the structure fits the

crystallographic data). The strain energy of the P^{1+} state in the original crystal structure is extremely high, 1323 kJ/mol, indicating that the structure is totally unrealistic and deviates severely from the optimum QM structure. In a good structure, the strain energy should be low: In the high-resolution 3U7Q crystal structure, the strain energy of the P-cluster is 67–147 kJ/mol for the P^{N} and P^{2+} states, and 10–24 kJ/mol after quantum refinement. For our best quantum-refined structure for 6CDK, the strain energy is only 6 kJ/mol and the RSZD scores indicate that the quantum-refined structure fits better to the X-ray data than the original structure, suggesting a significant improvement of the P-cluster.

The success of the two applications shows that the new ComQumX-2QM approach can be used to sort out what is actually seen in disordered crystal structures and to give locally improved coordinates that are in accordance with both the crystallographic raw data and QM calculations.

4.10. Paper X

As mentioned in Section 3, there are three types of the nitrogenase, differing in the nature of the metal ion in the active site. Molybdenum nitrogenase contains a molybdenum ion in the active site, which has been extensively studied[3,4] and is also the main subject of this thesis. However, there also exist alternative nitrogenases, in which the Mo ion is replaced by either vanadium or iron. The crystal structure of V-nitrogenase is known, showing that the active-site FeV cluster is similar to FeMo cluster, except that one of the belt sulfide ions (S3A) is replaced by a carbonate or nitrate ion.[167]

In 2018, a 1.2 Å crystal structure of vanadium nitrogenase was reported by Einsle and coworkers,[99] showing that a belt sulfide ion (S2B) is displaced from the cluster, leaving a vacancy occupied by a light atom. This light atom was interpreted as a NH^{2-} ligand based on an analysis of the hydrogen bonds and electron density. Consequently, the crystal was claimed to show the E_6 state in the reaction mechanism. Moreover, it was found that the side chain of a glutamine residue (Gln-176) had rotated towards the FeV cluster and appeared in a different conformation than in the resting-state of FeV cluster[167] and of all structures of FeMo cluster. They also observed a new density 7 Å from the FeV cluster in a cavity created by to the reorientation of Gln-176. It could be modelled as SH^-, so they suggested that this represents a storage site for the dissociated S2B ligand. However, soon after the publication, Bjornsson performed QM/MM calculations and showed that the crystal structure is better reproduced if the light atom is modelled as a bridging OH^- group, rather than a NH^{2-} group.[168]

QM/MM calculations do not employ any crystallographic data, besides the starting structure. Therefore, it is biased towards the starting structure. Quantum refinement is a more accurate approach, in that employs the crystallographic data during the geometry optimisations and therefore gives a structure that is an optimum compromise between crystallography and quantum mechanics. Moreover, it allows us to decide which structure fits the crystal structure best using crystallographic measures, rather than looking at distances, which are biased by the current model. Therefore, we employed quantum refinement in Paper X to better distinguish between the O and N ligands. We compared N^{3-}, NH^{2-} and NH_2^-, which corresponds to the E_5, E_6 and E_7 states, OH^- and O^{2-} for the resting state. We also considered other oxidation levels and spin states of the cluster, and different protonation state of His-180 (which can form a hydrogen bond to the ligand). We evaluated the refinements based on the absolute maximum RSZD scores of the ligand, Gln-176 and His-180, as well as the strain energy of the QM systems.

The results showed that all models gave similar RSZD values for His-180 (around 2.0–2.2) and a restricted variation for Gln-176 (8.9–10.8), whereas there were

vast differences between various ligands (9.7–22.1). The best results were obtained for an OH^- ligand (9.7–10.0), followed by NH_2^- (12.4), whereas the other three ligands gave worse results (16–22).

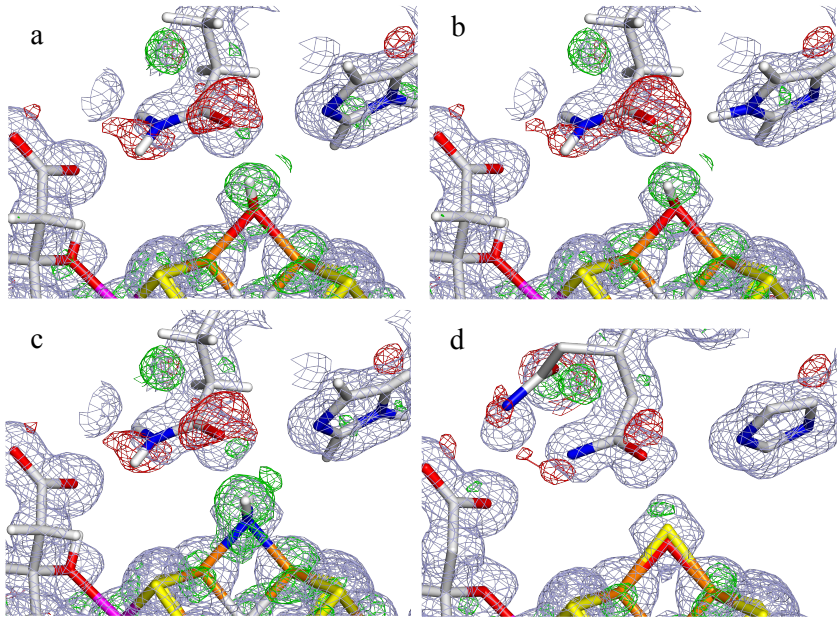


Figure 4.14: Electron-density maps of the best quantum-refined models of the 6FEA crystal structure: a) OH—HID, b) OH—HIE and c) NH_2^-–HID. d) shows a structure with both OH^- (0.83 occupancy) and S2B (0.17 occupancy) and two conformations of Gln-176 (0.89 occupancy of the flipped conformation, employed in the other structures, and 0.11 occupancy of the non-flipped conformation). The $2mF_o - DF_c$ maps are contoured at 1.0 σ (blue) and the $mF_o - DF_c$ maps are contoured at +3.0 σ (green) and –3.0 σ (red).

However, even for the best three structures (OH^-–HID, OH^-–HIE and NH_2^-–HID), the RSZD scores are much larger than 3, which is normally considered as the limit of an acceptable structure. This can also be seen from the electron-density maps in Figure 4.14. There is a large positive difference density blob around the OH^-/NH_2^- ligand, which may indicate that the heavier S2B ion is not fully dissociated. This is also supported by a large negative difference density blob around Gln-176, which indicates that the rotated conformation is not fully occupied. This was actually also noted by the original crystallographers, suggesting that there was <5% of the original structure present.[99] This hypothesis was confirmed by a standard crystallographic occupancy refinement employing both OH^- and S2B in the bridging position, as well as Gln-176 in both the rotated and the original (non-rotated as in the FeMo cluster) conformations. This led to occupancies of 0.17 for S2B and 0.83 for OH^-. Likewise, the occupancies for Gln-176 became 0.89 for the rotated and 0.11 for the original conformation. This led to strongly decreased RSZD values for the ligand and the

two residues, 2.0–2.6. The improvement is further illustrated by the electron difference density map (Figure 4.14d), showing appreciably smaller difference-density blobs around both the bridging ligand and Gln-176.

Consequently, we repeated the quantum-refinement calculations with two conformations of the QM system (i.e. using ComQumX-2QM from Paper IX), but employing the same occupancy for the ligand (again N^{3-}, NH^{2-}, NH_2^-, OH^- or O^{2-}) and Gln-176 (0.86/0.14). The results showed that OH^- still gave a lower RSZD score (1.1–1.2) than the other ligands (3.0–5.4). The HID conformation of His-180 gave a lower RSZD score for Gln-176 (5.7, compared to 7.3), but an appreciably higher strain energy (165, compared to 128 kJ/mol) and also a 104 kJ/mol higher QM energy. Therefore, OH^-–HIE seems to be the best interpretation of the crystal structure, together with a 0.14 portion of a dissociated S2B and a non-rotated Gln-176. This structure is shown in Figure 4.15.

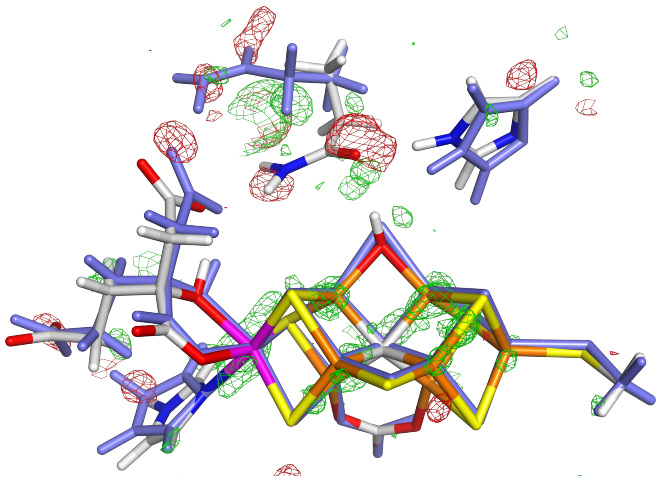


Figure 4.15: Electron-density difference maps of the best ComQumX-2QM quantum-refined OH^-–HIE model of the 6FEA crystal structure. Two conformations were used for the QM system. The first (with 86% occupancy, shown in atomic colours) represents the conformation reported in the crystal structure, involving an unknown ligand replacing S2B to the storage site and Gln-176 in the flipped conformation. The other conformation (14% occupancy, shown in slate) represents a normal E_0 resting state with S2B bound to the cluster and Gln-176 in a non-flipped conformation.

5. Conclusions and Outlook

The aim of this thesis has been to get a better understanding of the function and mechanism of nitrogenase based on computational methods with incorporated experimental information. To this end, we have extensively used QM/MM and quantum refinement to improve atomic structures, reveal the electronic states, find intermediates and predict a reaction mechanism.

Nitrogenase is one of the most complicated enzymes in nature.[1–3] It is a large protein and it contains two very complicated metal clusters, each with eight metal ions. This makes QM calculations with DFT methods very challenging. The prime problem with nitrogenase is the large number of possibilities for all reactions and states. We have identified over 50 possible binding sites for both protons and N_2 in the FeMo cluster. Moreover, each state has at least 35 different BS states. This gives an extremely large number of possibilities to test and most previous investigations have only tested a tiny fraction of these, typically without explaining which states were selected. Our starting hypothesis was that this was one important reason for the diverging results of previous computational investigations.[4] Therefore, we decided to test as many structures as possible and use well-defined and systematic heuristic procedures to decide what structures to test. With such an approach, it is necessary to do the investigations in many small steps.

Since we know the protein environment is important for computational characterisation of the active site, we decided to take the whole protein into account in all calculations, using the QM/MM method. At the start of our work, only a single QM/MM study had been published of nitrogenase,[113] which did not include the whole enzyme. Therefore, we decided to set up the calculations from scratch and to include the entire tetramer in the calculations, since there is no natural way to divide it into smaller pieces. As we knew that we would do many studies based on this set up, we spent much time on it, systematically deciding the protonation state of all amino acid residues. In particular, the protonation states of eight key residues around the active site could not be unambiguously predicted from standard protonation-state assignment, including the crucial homocitrate ligand bound to the Mo, which may directly affect the electronic structure of the cluster. Therefore, we performed MD simulations of the whole protein and used RMSD values as the assessment criteria to derive plausible protonation states. However, for the homocitrate ligand, a more

convincing conclusion was obtained based on the quantum-refinement calculations.

With the well-prepared system from Paper I, we next had to design a procedure to deal with the BS states. This was done in Paper II. We first investigated how the relative energy of the BS states was affected by the DFT method, the basis sets, the surrounding protein, as well as the protonation and the oxidation state of the cluster. Based on this, we proposed a practical procedure to deal with the BS states in the following studies.

With these initial studies, we could then start the real investigation of the reaction mechanism of nitrogenase. The aim was to do an exhaustive systematic study of all possible cases, while keeping the number of calculations manageable. Thus, in Paper III, we studied the structure (i.e. the protonation sites) of the four first intermediates in the reaction mechanism, E_1–E_4. To make such a study meaningful, we first needed to settle that the resting is not protonated, which was done by quantum refinement. Then, we systematically investigated all possible protonation states at each E_n level, starting from the best structure for the previous oxidation level. Unfortunately, it turned out that for the E_2–E_4 states, different DFT methods gave different predictions: The TPSS functional predicted protonation of the Fe ions (in the form of hydride ions), whereas the B3LYP functional strongly preferred protonation of the central carbide ion. Although this observation explained the controversy between Siegbahn and the Hoffman group,[117,155] it is of course a large problem in the study of nitrogenase.

Therefore, we studied the problem further in Paper IV, including 13 different DFT methods. The calculations showed an unprecedented large difference in the relative energies for the various protonation states (up to almost 1100 kJ/mol) obtained by the different methods. Hybrid functionals preferred carbide protonation, whereas the non-hybrid functionals favoured hydride ions. The problem was especially large when the formal oxidation state of the cluster changes or when the number of Fe–C or Fe–S bonds change. We made a first attempt to decide which method gives the most reliable results, but this was not conclusive: GGA functionals and TPSSh give the best geometries of the resting state, but B3LYP and PBE0 seem to give more reliable H_2 dissociation energies. On the other hand, B3LYP gives an incorrect electronic state. Currently, the best approach seems to be to always do all studies with one pure and one hybrid functional, to check whether the results are sensitive to the DFT method.

The results in both Paper III and IV indicated that no DFT method suggested that the best E_4 state involves two bridging hydride ions, as proposed by ENDOR measurements.[3,88,89] Therefore, we decided to systematically investigate such structures in Paper V. We first determined the best positions of the two protonated sulfide ions and then calculated the relative energies of all possible pairs of bridging hydride ions. The study showed that the best structures always involved

hydride bridges between the Fe2/6 and Fe3/7 pairs. These are the same as suggested by Hoffman and coworkers,[3,89,169] but our calculations indicated that it was more favourable if the Fe2/6 bridge is on the other side of the S2B ion and if the protons on S2B and S5A points in other directions. However, most importantly, this study showed that these double-hydride-bridged structures actually are the best possible E_4 structures with pure functionals. An important reason for this changed conclusion (compared to Papers III and IV) is that we found a new type of BS states, involving only two Fe ions with minority spin.

In Paper VI, we studied the binding of N_2 to the FeMo cluster. Since the structure of E_4 is still controversial and involve so many possible positions of the protons, we decided to instead study the binding of N_2H_2 to the E_0 state, i.e. after the dissociation of H_2 from E_4 and protonation of N_2 by the remaining two protons. Thereby, the number of possible binding modes was reduced to ~60, including binding of trans- and cis-HNNH, as well as NNH_2, and including end-on and side-on binding modes to one, two or four Fe ions. We found that the best structure had trans-HNNH binding terminally to Fe2. However, several other structures were rather close in energy.

In Paper VII, we studied the last reactions of nitrogenase, starting from the N_2-bound state and going through the E_5–E_8 states to form two ammonia molecules. To avoid the problem of the binding site of N_2 (the study was started before Paper VI was finished), we studied nitrogenase with the S2B ligand dissociated, inspired by two recent crystal structures indicating that such a dissociation is reversible and might be involved in the reaction mechanism.[70,99] Our study showed that such a reaction is possible and thermodynamically favourable. We systematically investigated all possible protonation states and binding modes, as well as the protonation state of His-195, which can form hydrogen bonds to the substrate. Our study suggested that the reaction involves NNH_2 binding end-on to both Fe2 and Fe6, H_2NNH in a mixed binding mode (one N binds to both Fe2 and Fe6, the other to only Fe6), as well as H_2NNH_2, NH_2 and NH_3 all bridging Fe2 and Fe6. Thus, the mechanism is mainly alternating, even if the first intermediate (NNH_2) normally is connected to a distal mechanism.

In Paper VIII, we turned to the P-cluster of nitrogenase. We first decided the proper BS states of this cluster in the four most reduced oxidation states. Again, this led to some problems with diverging results of the TPSS and B3LYP methods. Next, we settled the protonation states of the four oxidation states. In particular, we showed with both QM/MM and quantum-refinement calculations that the one-electron oxidised state (P^{1+}) involves a protonated backbone N atom of Cys-C88, but a deprotonated Ser-D188 residue. This study opens up for future investigations of the interplay between the P-cluster and the FeMo cluster.

The quantum-refinement studies in Paper VIII indicated that the crystal structure of P^{1+} involves a mixture of the P^{1+} and P^{2+} states. To be able to treat such

mixtures with quantum refinement, we developed in Paper IX a new quantum-refinement approach that allows for disorder in the quantum system, by simply performing two separate QM calculations of each state. We showed that the method works properly by re-refining the P-cluster in two crystal structures of nitrogenase,[47,48] one involving P^N and P^{2+} and the other involving P^{1+} and P^{2+}. The two structures also have widely different resolutions. This study clearly showed that our protonation assignment of P^{1+} is correct and that the P^{1+} structure suggested by the crystallographers involves several strongly dubious Fe–S and Fe–Fe distances.

Finally, in Paper X, we exploited this new quantum-refinement approach to interpret the crystal structure of V-nitrogenase with a putative E_6 or E_7 reaction intermediate.[99] With standard quantum refinement we first confirmed the results of Bjornsson and coworkers[168] that the structure rather involves an OH^--inhibited state. However, even with this interpretation there was significant features in the electron-density difference maps. We showed that these can be interpreted as a significant amount of undissociated S2B. Therefore, we repeated the calculations with our new quantum-refinement approach, showing that the structure indeed contains a OH^- ligand.

In conclusion, we have in this thesis developed a systematic QM/MM and quantum-refinement based approach to study the reaction mechanism of nitrogenase. We have then used this approach to study several interesting aspects of the nitrogenase reaction. Even if the study has shown that the problems involving BS states and the DFT functional are often much worse than is normally assumed, several interesting results are obtained. In particular, we start to see a convergence between the results obtained with computational and experimental studies.

Yet, even if we have taken many significant steps towards the understanding of the reaction mechanism of nitrogenase, much remains to do in future studies. As long as there is no experimental consensus whether the S2B ligand dissociates or remains bound during catalysis, this needs to be considered as two reaction scenarios, which should be examined with separate calculations. Therefore, for Mo-nitrogenase, with the bridging S2B sulfide ion remaining bound to the cluster, we would like to:

- Extend the study of N_2H_2 binding in Paper VI to the remaining E_5–E_8 state (like in Paper VII).
- Study the binding of N_2 to the FeMo cluster, together with the reductive elimination of H_2. This can be done by working forward from the E_4 structure in Paper V or working backward from the N_2H_2 structures in Paper VI.

Likewise, for Mo-nitrogenase with a dissociated S2B sulfide ion, we would like to:

- Complement the study in Paper VII with an investigation of possible proton transfer paths from the surrounding environment to the active site together with activation energies.
- Study the binding of N_2 and the reductive elimination of H_2 in the same way as when S2B remains bound.
- Study the actual dissociation reaction of S2B from the cluster. At what E_n state does it take place? Is it associated with the binding of N_2?

With these studies, it would be possible to compare the two reaction mechanisms and predict if and how the S2B dissociation may boost the N_2 reduction. In this respect, it may be necessary to also investigate systems with the S2B ion half-dissociated, as recently has been suggested by two groups.[162,163]

References

[1] B. K. Burgess, D. J. Lowe, *Chem. Rev.* **1996**, *96*, 2983–3012.

[2] B. Schmid, H.-J. Chiu, V. Ramakrishnan, J. B. Howard, D. C. Rees, in *Handb. Met.*, John Wiley & Sons, Ltd, **2006**, pp. 1025–1036.

[3] B. M. Hoffman, D. Lukoyanov, Z.-Y. Yang, D. R. Dean, L. C. Seefeldt, *Chem. Rev.* **2014**, *114*, 4041–4062.

[4] I. Dance, *ChemBioChem* **2020**, DOI: 10.1002/cbic.201900636.

[5] O. Einsle, *J. Biol. Inorg. Chem.* **2014**, *19*, 737–745.

[6] P. E. Smith, B. M. Pettitt, *J. Phys. Chem.* **1994**, *98*, 9700–9711.

[7] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864–B871.

[8] T. Ziegler, *Can. J. Chem.* **1995**, *73*, 743–761.

[9] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133.

[10] S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200–1211.

[11] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098–3100.

[12] J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822–24.

[13] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.

[14] J. P. Perdew, K. Burke, M. Ernzerhof, B. Kieron, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[15] S. Grimme, *J. Comput. Chem.* **2006**, *27*, 1787–1799.

[16] Y. Zhao, D. G. Truhlar, *J. Chem. Phys.* **2006**, *125*, 194101.

[17] V. N. Staroverov, G. E. Scuseria, J. Tao, J. P. Perdew, *J. Chem. Phys.* **2003**, *119*, 12129–12137.

[18] J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, *Phys. Rev. Lett.* **2003**, *91*, 146401.

[19] J. P. Perdew, M. Ernzerhof, K. Burke, *J. Chem. Phys.* **1996**, *105*, 9982.

[20] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[21] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215–241.

[22] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 1372.

[23] Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **2006**, *110*, 13126–13130.

[24] F. Neese, *J. Biol. Inorg. Chem.* **2006**, *11*, 702–711.

[25] A. R. Leach, *Molecular Modelling: Principles and Applications*, **2001**.

[26] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

[27] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Comput. Chem* **2004**, *25*, 1157–1174.

[28] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.

[29] C. I. Bayly, P. Cieplak, W. D. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269–10280.

[30] F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, Ltd, Chichester, **2017**.

[31] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.

[32] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227–249.

[33] H. M. Senn, W. Thiel, *Angew. Chemie - Int. Ed.* **2009**, *48*, 1198–1229.

[34] U. Ryde, *Methods Enzymol.* **2016**, *577*, 119–158.

[35] N. Reuter, A. Dejaegere, B. Maigret, M. Karplus, *J. Phys. Chem. A* **2000**, *104*, 1720–1735.

[36] U. Ryde, *J. Comput. Aided. Mol. Des.* **1996**, *10*, 153–164.

[37] L. Hu, P. Söderhjelm, U. Ryde, *J. Chem. Theory Comput.* **2011**, *7*, 761–777.

[38] J. Gao, P. Amara, C. Alhambra, M. J. Field, *J. Phys. Chem. A* **1998**, *102*, 4714–4721.

[39] M. R. A. Blomberg, T. Borowski, F. Himo, R.-Z. Liao, P. E. M. Siegbahn, *Chem. Rev.* **2014**, *114*, 3601–58.

[40] L. Hu, J. Eliasson, J. Heimdal, U. Ryde, *J. Phys. Chem. A* **2009**, *113*, 11793–11800.

[41] S. Sumner, P. Söderhjelm, U. Ryde, *J. Chem. Theory Comput.* **2013**, *9*, 4205–4214.

[42] C. V. Sumowski, C. Ochsenfeld, *J. Phys. Chem. A* **2009**, *113*, 11734–11741.

[43] S. Roßbach, C. Ochsenfeld, *J. Chem. Theory Comput.* **2017**, *13*, 1102–1107.

[44] R. Z. Liao, W. Thiel, *J. Comput. Chem.* **2013**, *34*, 2389–2397.

[45] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235–242.

[46] G. Rhodes, *Crystallography Made Crystal Clear*, Academic Press, New York, NY, **2006**.

[47] S. M. Keable, O. A. Zadvornyy, L. E. Johnson, B. Ginovska, A. J. Rasmussen, K. Danyal, B. J. Eilers, G. A. Prussia, A. X. LeVan, S. Raugei, et al., *J. Biol. Chem.* **2018**, *293*, 9629–9635.

[48] T. Spatzal, M. Aksoyoglu, L. Zhang, S. L. a. Andrade, E. Schleicher, S. Weber, D. C. Rees, O. Einsle, *Science (80-.).* **2011**, *334*, 940–940.

[49] R. J. Read, *Acta Crystallogr. Sect. A* **1986**, *42*, 140–149.

[50] A. T. Brünger, *Nature* **1992**, *355*, 472–475.

[51] I. J. Tickle, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, *68*, 454–467.

[52] N. S. Pannu, R. J. Read, *Acta Crystallogr. Sect. A Found. Crystallogr.* **1996**, *A52*,

659–668.

[53] P. D. Adams, N. S. Pannu, R. J. Read, A. T. Brünger, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 5018–5023.

[54] G. J. Kleywegt, T. A. Jones, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1998**, *54*, 1119–1131.

[55] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, et al., *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 213–221.

[56] G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr. Sect. D* **1997**, *53*, 240–255.

[57] O. S. Smart, T. O. Womack, C. Flensburg, P. Keller, W. Paciorek, A. Sharff, C. Vonrhein, G. Bricogne, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, *68*, 368–380.

[58] U. Ryde, K. Nilsson, *J. Am. Chem. Soc.* **2003**, *125*, 14232–14233.

[59] U. Ryde, L. Olsen, K. Nilsson, *J. Comput. Chem.* **2002**, *23*, 1058–1070.

[60] R. A. Engh, R. Huber, *Acta Crystallogr. Sect. A* **1991**, *47*, 392–400.

[61] F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, F. Weigend, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 91–100.

[62] A. T. Brunger, P. D. Adams, G. M. Clore, W. L. Delano, P. Gros, R. W. Grosse-kunstleve, J. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, et al., *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1998**, *54*, 905–921.

[63] A. T. Brunger, *Nat. Protoc.* **2007**, *2*, 2728–33.

[64] B. E. Smith, *Science (80-.).* **2002**, *297*, 1654–1655.

[65] K. M. Lancaster, M. Roemelt, P. Ettenhuber, Y. Hu, M. W. Ribbe, F. Neese, U. Bergmann, S. DeBeer, *Science (80-.).* **2011**, *334*, 974–977.

[66] R. Bjornsson, F. A. Lima, T. Spatzal, T. Weyhermüller, P. Glatzel, E. Bill, O. Einsle, F. Neese, S. DeBeer, *Chem. Sci.* **2014**, *5*, 3096–3103.

[67] L. E. Roth, F. A. Tezcan, *Methods Mol. Biol.* **2011**, *766*, 147–164.

[68] J. Kim, D. C. Rees, *Science (80-.).* **1992**, *257*, 1677–1682.

[69] O. Einsle, F. A. Tezcan, S. L. A. Andrade, B. Schmid, M. Yoshida, J. B. Howard, D. C. Rees, *Science (80-.).* **2002**, *297*, 1696.

[70] T. Spatzal, K. a Perez, O. Einsle, J. B. Howard, D. C. Rees, *Science (80-.).* **2014**, *345*, 1620–1623.

[71] R. R. Eady, *Chem. Rev.* **1996**, *96*, 3013–3030.

[72] H. Schindelin, C. Kisker, J. L. Schlessman, J. B. Howard, D. C. Rees, *Nature* **1997**, *387*, 370.

[73] J. Rawlings, V. K. Shah, J. R. Chisnell, W. J. Brill, R. Zimmermann, E. Münck, W. H. Orme-Johnson, *J. Biol. Chem.* **1978**, *253*, 1001–4.

[74] H. I. Liu, B. K. Burgess, C. R. Natoli, A. Filipponi, N. Gavini, B. Hedman, A. Di Cicco, K. O. Hodgson, *J. Am. Chem. Soc.* **1994**, *116*, 2418–2423.

[75] S. P. Cramer, K. O. Hodgson, W. O. Gillum, L. E. Mortenson, *J. Am. Chem. Soc.*

1978, *100*, 3398–3407.

[76] R. A. Venters, M. J. Nelson, P. A. McLean, A. E. True, M. A. Levy, B. M. Hoffman, W. H. Orme-Johnson, *J. Am. Chem. Soc.* **1986**, *108*, 3487–3498.

[77] B. Hedman, P. Frank, S. F. Gheller, A. L. Roe, W. E. Newton, K. O. Hodgson, *J. Am. Chem. Soc.* **1988**, *110*, 3798–3805.

[78] T. V Harris, R. K. Szilagyi, *Inorg. Chem* **2011**, *50*, 4811–4824.

[79] H.-I. Lee, B. J. Hales, B. M. Hoffman, *J. Am. Chem. Soc.* **1997**, *119*, 11395–11400.

[80] S. J. Yoo, H. C. Angove, V. Papaefthymiou, B. K. Burgess, E. Münck, *J. Am. Chem. Soc.* **2000**, *122*, 4926–4936.

[81] T. Spatzal, J. Schlesier, E.-M. Burger, D. Sippel, L. Zhang, S. L. A. Andrade, D. C. Rees, O. Einsle, *Nat. Commun.* **2016**, *7*, 10902.

[82] R. Bjornsson, F. Neese, S. DeBeer, *Inorg. Chem.* **2017**, *56*, 1470–1477.

[83] L. C. Seefeldt, B. M. Hoffman, D. R. Dean, *Annu. Rev. Biochem.* **2009**, *78*, 701–722.

[84] R. N. F. Thorneley, D. J. Lowe, in *Molybdenum Enzym.* (Ed.: T.G. Spiro), Wiley, New York, **1985**, pp. 221–284.

[85] C.-H. Kim, W. E. Newton, D. R. Dean, *Biochemistry* **1995**, *34*, 2798–2808.

[86] M. J. Dilworth, K. Fisher, C. H. Kim, W. E. Newton, *Biochemistry* **1998**, *37*, 17495–505.

[87] C. Van Stappen, A. T. Thorhallsson, L. Decamps, R. Bjornsson, S. DeBeer, *Chem. Sci.* **2019**, *10*, 9807–9821.

[88] R. Y. Igarashi, M. Laryukhin, P. C. Dos Santos, H.-I. Lee, D. R. Dean, L. C. Seefeldt, B. M. Hoffman, *J. Am. Chem. Soc.* **2005**, *127*, 6231–6241.

[89] V. Hoeke, L. Tociu, D. A. Case, L. C. Seefeldt, S. Raguei, B. M. Hoffman, *J. Am. Chem. Soc.* **2019**, *141*, 11984–11996.

[90] D. Lukoyanov, Z.-Y. Yang, N. Khadka, D. R. Dean, L. C. Seefeldt, B. M. Hoffman, *J. Am. Chem. Soc.* **2015**, *137*, 3610–3615.

[91] D. Lukoyanov, N. Khadka, Z.-Y. Yang, D. R. Dean, L. C. Seefeldt, B. M. Hoffman, *J. Am. Chem. Soc.* **2016**, *138*, 10674–10683.

[92] S. Raugei, L. C. Seefeldt, B. M. Hoffman, *Proc. Natl. Acad. Sci.* **2018**, *115*, 10521–10530.

[93] J. Chatt, *Annu. Proc. Phytochem. Soc. Eur.* **1980**, *18*, 1–18.

[94] J. Chatt, J. R. Dilworth, R. L. Richards, *Chem. Rev.* **1978**, *78*, 589–625.

[95] D. V Yandulov, R. R. Schrock, *Science (80-.).* **2003**, *301*, 76–78.

[96] R. R. Schrock, *Acc. Chem. Res.* **2005**, *38*, 955–962.

[97] R. R. Schrock, *Angew. Chemie Int. Ed.* **2008**, *47*, 5512–5522.

[98] J. B. Varley, Y. Wang, K. Chan, F. Studt, J. K. Nørskov, *Phys. Chem. Chem. Phys.* **2015**, *17*, 29541–29547.

[99] D. Sippel, M. Rohde, J. Netzer, C. Trncik, J. Gies, K. Grunau, I. Djurdjevic, L. Decamps, S. L. A. Andrade, O. Einsle, *Science (80-.).* **2018**, *359*, 1484–1489.

[100] D. Lukoyanov, S. A. Dikanov, Z.-Y. Yang, B. M. Barney, R. I. Samoilova, K. V Narasimhulu, D. R. Dean, L. C. Seefeldt, B. M. Hoffman, *J. Am. Chem. Soc.* **2011**, *133*, 11655–11664.

[101] R. N. F. Thornely, R. R. Eady, D. J. Lowe, *Nature* **1978**, *272*, 557–558.

[102] R. N. F. Thorneley, D. J. Lowe, *Biochem. J.* **1984**, *224*, 887–894.

[103] P. P. Hallmen, J. Kästner, *Zeitschrift fur Anorg. und Allg. Chemie* **2015**, *641*, 118–122.

[104] I. Dance, *Zeitschrift für Anorg. und Allg. Chemie* **2015**, *641*, 91–99.

[105] P. E. M. Siegbahn, *J. Am. Chem. Soc.* **2016**, *138*, 10485–10495.

[106] M. L. McKee, *J. Phys. Chem. A* **2016**, *120*, 754–764.

[107] L. Rao, X. Xu, C. Adamo, *ACS Catal.* **2016**, *6*, 1567–1577.

[108] F. Tuczek, in *RSC Met. Ser. 7* (Eds.: R. Hille, C. Schulzke, M.L. Kirk), Royal Society Of Chemistry, Cambridge, **2017**, pp. 223–274.

[109] I. Dance, *J. Biol. Inorg. Chem.* **1996**, *1*, 581–586.

[110] K. K. Stavrev, M. C. Zerner, *Int. J. Quantum Chem.* **1998**, *70*, 1159–1168.

[111] P. E. M. Siegbahn, J. Westerberg, M. Svensson, R. H. Crabtree, *J. Phys. Chem. B* **1998**, *102*, 1615–1623.

[112] T. Lovell, J. Li, T. Liu, D. A. Case, L. Noodleman, *J. Am. Chem. Soc.* **2001**, *123*, 12392–12410.

[113] H. Xie, R. Wu, Z. Zhou, Z. Cao, *J. Phys. Chem. B* **2008**, *112*, 11435–11439.

[114] J. Kästner, P. E. Blöchl, *J. Am. Chem. Soc.* **2007**, *129*, 2998–3006.

[115] B. Benediktsson, R. Bjornsson, *Inorg. Chem.* **2017**, *56*, 13417–13429.

[116] I. Dance, *Inorganics* **2019**, *7*, DOI: 10.3390/inorganics7010008.

[117] P. E. M. Siegbahn, *J. Comput. Chem.* **2018**, *39*, 743–747.

[118] P. E. M. Siegbahn, *Phys. Chem. Chem. Phys.* **2019**, *21*, 15747–15759.

[119] P. E. M. Siegbahn, *Inorg. Chem.* **2018**, *57*, 1090–1095.

[120] Z. Li, J. Li, N. S. Dattani, C. J. Umrigar, G. K.-L. Chan, *J. Chem. Phys.* **2019**, *150*, 24302.

[121] S. Sharma, K. Sivalingam, F. Neese, C. Kin-Lic, *Nat Chem* **2014**, *6*, 927–933.

[122] L. Noodleman, *J. Chem. Phys.* **1981**, *74*, 5737–5743.

[123] L. Cao, U. Ryde, *Int. J. Quantum Chem.* **2018**, *118*, 1.

[124] D. Lukoyanov, V. Pelmenschikov, N. Maeser, M. Laryukhin, T. C. Yang, L. Noodleman, D. R. Dean, D. A. Case, L. C. Seefeldt, B. M. Hoffman, *Inorg. Chem.* **2007**, *46*, 11437–11449.

[125] I. Dance, *Inorg. Chem.* **2011**, *50*, 178–192.

[126] C. Greco, P. Fantucci, U. Ryde, L. de Gioia, *Int. J. Quantum Chem.* **2011**, *111*, 3949–3960.

[127] R. K. Szilagyi, M. A. Winslow, *J. Comput. Chem.* **2006**, *27*, 1385–1397.

[128] W. Kaim, B. Schwederski, *Bioinorganic Chemistry: Inoroganic Elements in the Chemistry of Life*, John Wiley & Sons, Inc., Chichester, **1991**.

[129] Z.-Y. Yang, K. Danyal, L. C. Seefeldt, in *Methods Mol. Biol.*, **2011**, pp. 267–292.

[130] A. J. Pierik, H. Wassink, H. Haaker, W. R. Hagen, *Eur. J. Biochem.* **1993**, *212*, 51–61.

[131] K. Rupnik, Y. Hu, C. C. Lee, J. A. Wiig, M. W. Ribbe, B. J. Hales, *J. Am. Chem. Soc.* **2012**, *134*, 13749–13754.

[132] C. P. Owens, F. E. H. Katz, C. H. Carter, V. F. Oswald, F. A. Tezcan, *J. Am. Chem. Soc.* **2016**, *138*, 10124–10127.

[133] W. N. Lanzilotta, J. Christiansen, D. R. Dean, L. C. Seefeldt, *Biochemistry* **1998**, *37*, 11376–11384.

[134] J. M. Chan, J. Christiansen, D. R. Dean, L. C. Seefeldt, *Biochemistry* **1999**, *38*, 5779–5785.

[135] K. Danyal, D. R. Dean, B. M. Hoffman, L. C. Seefeldt, *Biochemistry* **2011**, *50*, 9255–9263.

[136] M. Rohde, D. Sippel, C. Trncik, S. L. A. Andrade, O. Einsle, *Biochemistry* **2018**, *57*, 5497–5504.

[137] J. M. Mouesca, L. Noodleman, D. A. Case, *Inorg. Chem.* **1994**, *33*, 4819–4830.

[138] Z. Li, S. Guo, Q. Sun, G. K.-L. Chan, *Nat. Chem.* **2019**, *11*, 1026–1033.

[139] R. C. Tittsworth, B. J. Hales, *J. Am. Chem. Soc.* **1993**, *115*, 9763–9767.

[140] H. C. Angove, S. J. Yoo, B. K. Burgess, E. Münck, *J. Am. Chem. Soc.* **1997**, *119*, 8730–8731.

[141] B. B. Wenke, T. Spatzal, D. C. Rees, *Angew. Chemie Int. Ed.* **2019**, *58*, 3894–3897.

[142] T. H. Rod, U. Ryde, *J. Chem. Theory Comput.* **2005**, *1*, 1240–1251.

[143] M. Kaukonen, P. Söderhjelm, J. Heimdal, U. Ryde, *J. Phys. Chem. B* **2008**, *112*, 12537–12548.

[144] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*, 525–537.

[145] *Schrödinger Release 2015-4: Maestro*, Schrödinger, LLC, New York, NY, **2016**.

[146] T. H. Rod, U. Ryde, *Phys. Rev. Lett.* **2005**, *94*, 1–4.

[147] A. Warshel, A. Dryga, *Proteins Struct. Funct. Bioinforma.* **2011**, *79*, 3469–3484.

[148] E. Alexov, E. L. Mehler, N. Baker, A. M. Baptista, Y. Huang, F. Milletti, J. Erik Nielsen, D. Farrell, T. Carstensen, M. H. M. Olsson, et al., *Proteins Struct. Funct. Bioinforma.* **2011**, *79*, 3260–3275.

[149] J. Cheng, X. Liu, J. VandeVondele, M. Sulpizi, M. Sprik, *Acc. Chem. Res.* **2014**, *47*, 3522–3529.

[150] T. Lovell, J. Li, D. A. Case, L. Noodleman, *J. Biol. Inorg. Chem.* **2002**, *7*, 735–749.

[151] A. Schäfer, H. Horn, R. Ahlrichs, *J. Chem. Phys.* **1992**, *97*, 2571–2577.

[152] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–305.

[153] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.

[154] J. Li, R. A. Mata, U. Ryde, *J. Chem. Theory Comput.* **2013**, *9*, 1799–1807.

[155] D. Lukoyanov, N. Khadka, D. R. Dean, S. Raugei, L. C. Seefeldt, B. M. Hoffman, *Inorg. Chem.* **2017**, *56*, 2233–2240.

[156] L. Cao, O. Caldararu, U. Ryde, *J. Chem. Theory Comput.* **2018**, *14*, 6653–6678.

[157] L. Cao, U. Ryde, *Phys. Chem. Chem. Phys.* **2019**, in press; DOI 10.1039/C9CP06930A.

[158] I. Dance, *Dalt. Trans.* **2008**, *2*, 5977–5991.

[159] I. Dance, *Dalt. Trans.* **2012**, *41*, 4859.

[160] B. M. Hoffman, D. Lukoyanov, D. R. Dean, L. C. Seefeldt, *Acc. Chem. Res.* **2013**, *46*, 587–595.

[161] D. Lukoyanov, B. M. Barney, D. R. Dean, L. C. Seefeldt, B. M. Hoffman, *Proc. Natl. Acad. Sci.* **2007**, *104*, 1451–1455.

[162] A. T. Thorhallsson, B. Benediktsson, R. Bjornsson, *Chem. Sci*. **2019**, *10*, 11110–11124.

[163] I. Dance, *Dalt. Trans.* **2019**, *48*, 1251–1262.

[164] P. Söderhjelm, U. Ryde, *J. Mol. Struct. THEOCHEM* **2006**, *770*, 199–219.

[165] L. Rulíšek, U. Ryde, *J. Phys. Chem. B* **2006**, *110*, 11511–11518.

[166] K. Nilsson, H.-P. Hersleth, T. H. Rod, K. K. Andersson, U. Ryde, *Biophys. J.* **2004**, *87*, 3437–47.

[167] D. Sippel, O. Einsle, *Nat. Chem. Biol.* **2017**, *13*, 956–960.

[168] B. Benediktsson, A. T. Thorhallsson, R. Bjornsson, *Chem. Commun.* **2018**, *54*, 7310–7313.

[169] R. Y. Igarashi, M. Laryukhin, P. C. Dos Santos, H.-I. Lee, D. R. Dean, L. C. Seefeldt, B. M. Hoffman, *J. Am. Chem. Soc.* **2005**, *127*, 6231–6241.

Acknowledgments

Eventually, I arrived to this part, a few hours before delivering the thesis to print. I have thought many times what I should write when reading other theses and also during my writing period. My decision was that I would not write anything here, but instead thank in private. Just a few seconds before, when I was waiting for my last figure of the thesis, memories reappear. Although the figure is rendered now, I still want to spend some time here.

Dear **Ulf**, my best supervisor, thank you for all the infinite support and endless patient to me. Without you, I wouldn't be the 'me' today. I still remember the first semester meeting in 2016, when you told me that I am not competitive. It was at that time, I realised how much I have lost myself. It was also at that time, I realised that science also need competitiveness, which explains our millions of results. Without you, I would not even dare to imagine that I can enjoy my PhD study so much. Thank you for allowing me to touch different fields and for leading me. From you, I learnt how passion can promote both life and work. Thank you for working with me and encourage me, a both (only sometimes) stubborn and picky me. I feel so sorry that I am not a person that is good at expressing gratitude by language, but you know who I am.

Esko, although you are not responsible for my studies, you are always so generous to help: To solve the problems I encountered, revision of the thesis and to offer me crystallography lessons. I feel so grateful when you tell me that I can contact you whenever I have problem.

Ebbe, for giving me good suggestions.

Thanks for all former members **Martin, Geng**, etc., and current members of our group: **Octav**, thank you for participating in my projects. It was a paper you selected for the journal club that gave me the idea to apply quantum refinement to homocitrate and started all the following cooperation. Thanks for teaching and helping me. **Magne**, thank you for your greetings. **Justin**, for solving my problems and sharing the cakes. **Joel**, for asking me 'how are you' when I began writing my thesis, for making fun of "djur" and for the nice chatting. **Vilhelm**, for cooperation and comfort. **Erik**, you told me that people are only interested in the acknowledgement part, which made me do not want to write it, until now when I am in such a hurry! Here comes the compliment: However, I would like to admit that you, the "cool" Erik, is my cool friend. You do not only supervise

me, explain quantum methods to me, but also chat with me so much, which helped me a lot. Dear **Majda**, thanks for teaching me how to perform FEP calculations. I enjoy all the time we spend together, go to yoga, have a cup of coffee, dumplings, go shopping and talk sooooo much. I will not forget the gifts you brought me and the food you made. I hope we will meet very soon and continue our dumplings and beer routine with Erik again!

Marie Skepö, thank you for helping me. **Mikael** and **Valera**, for all the technical support. **Jan**, for the suggestion of study plans. **Helena** and **Maria**, for the administrative help. **Per-Åke**, for helping me to fix the credits. A big thanks to everyone in the **Teokem** division, for all the greetings, fika and cake times.

Meanwhile, I would like to thank for **Junhao**, for the patience whenever I need help. **Hongduo** and **Feifei**, for all the fantastic Chinese food, excursions and card games. **Ruiyu**, for arranging parties and **Xiaoyan**, for nice chatting, encouragement and hanging out together.

Shaoran, this is the first time I call you so formally, instead of your nickname, Old pan. We had been the best friends for so many years. When I was in Beijing, you were in Japan, and when I am in Sweden, you are still in Japan. Nevertheless, I can still disturb you whenever I want, as you are so close to me. **Hongyuan**, when you tell me that I am living the life of your dreams, but that you are not at all jealous at me, instead imagining that you are me, I feel so sorry. On one hand, I understand you could not leave your kids to chase your own dream. On the other hand, I perhaps failed to live the life you expected. But with all your blessings, I will try to make it colourful and meaningful.

Thanks for **Paula**, for your professional skills on printing my thesis.

Thanks to all people that brought a lot of fun into my life. There are so many faces appearing in my mind but I wouldn't be able to list all your names here.

Finally, thanks to **my parents** for your endless love.

Nitrogenase is the only enzyme that can convert the inert nitrogen molecule to ammonia, so that it can be used in biosynthetic pathways. It contains a complicated acitve site, composed of eight metal ions, nine sulfur ions and one carbide ion (the FeMo cluster). Although it has been thoroughly studied with experimental and computational methods, the reaction mechanism is still not known and many conflicting hypotheses have been presented. To solve some of these problems, we have performed a thorough and systematic study of nitrogenase with various computational approaches. We have:

- Decided the protonation states of eight key amino acid residues around the active site and showed that the homocitrate ligand is singly prontated on the hydroxide group.
- Studied how the the broken-symmetry (BS) state for the FeMo cluster depends on the QM method, the basis sets, the surrounding protein and the protonation and oxidation state of the cluster.
- Predicted the most stable protonation state for the E_0–E_4 states of nitrogenase with two density functional theory (DFT) methods.
- Showed that different DFT methods give relative eneriges that can differ by almost 1100 kJ/mol for nitrogenase. This is the main reason for the diverging computational results.
- Showed that the most stable E_4 structure obtained with pure functionals has two hydride ions bridging between two pairs of iron ions, in agreement with experiments.
- Predicted that the most stable binding mode of N_2H_2 to nitrogenase involves trans-HNNH binding to Fe2.
- Suggested an alternating reaction mechanism for nitrogenase with a dissociated S2B ligand.
- Decided the most stable BS states for the P-cluster in four oxidation states and decided the protonation state of the one-electron oxidised state.
- Developed a novel quantum-refinement approach allowing for disorder in the QM system and applied it to the P-cluster in two crystal structures of nitrogenase.
- Shown by quantum refinement that a recent crystal structure of V-nitrogenase does not involve a N-derived ligand, but rather a hydride-inhibited state.

Printed by Media-Tryck, Lund 2020 NORDIC SWAN ECOLABEL 3041 0903

ISBN: 978-91-7422-741-3

Theoretical Chemistry
Department of Chemistry
Faculty of Science
Lund University