

Reaction mechanism of metalloenzymes studied by theoretical methods

Geng Dong



LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Science, Lund University, Sweden.
To be defended on 31st May 2018, at 9:00 in lecture hall B, Centre for
Chemistry and Chemical Engineering, Lund.

Faculty opponent

Prof. Maria João Ramos, University of Porto, Portugal

Grading committee

Dr. Marcus Lundberg (Uppsala University)

Prof. Christine McKenzie (University of Southern Denmark)

Assoc. Prof. Yaoquan Tu (Royal Institute of Technology)

Organization LUND UNIVERSITY Centre for Chemistry and Chemical Engineering P.O. Box 124, SE-221 00, Lund, Sweden.		Document name DOCTORAL DISSERTATION	
Author(s) Geng Dong		Date of issue 2018-05-04	
		Sponsoring organization	
Title and subtitle Reaction mechanism of metalloenzymes studied by theoretical methods			
Abstract Metalloenzymes catalyse a wide variety of reactions in nature. In the thesis, I have studied the reaction mechanism of three metalloenzymes, viz. [NiFe] hydrogenase (H ₂ ase), dimethyl sulfoxide reductase (DMSOR) and formate dehydrogenase (FDH), by theoretical methods, namely quantum mechanics (QM), combined quantum mechanical and molecular mechanics (QM/MM), as well as QM/MM thermodynamic cycle perturbation (QTCP). For H ₂ ase, we have studied the protonation states of the four cysteine residues in the active site at four intermediate states, the H ₂ binding site and the full reaction mechanism. Our results demonstrate that the Cys546 residue is most easily protonated by 14–51 kJ/mol, H ₂ binding to Ni ion in singlet state is most favourable by at least 47 kJ/mol, and the Ni-L state is not involved in the reaction mechanism. For the H ₂ binding, we have calibrated density-functional methods with advanced QM methods, like CCSD(T), DMRG-CASPT2 and CAS-srDFT. For DMSOR, we have studied the effect of the protein ligand in reaction mechanism. Our results indicate that enzymes with ligand with a single negative charge (serine, cysteine, selenocysteine, SH ⁻ and OH ⁻) are predicted to have two-step reaction mechanisms, giving an activation energy of 69–85 kJ/mol. However, the O ²⁻ and S ²⁻ ligands gave much higher activation energies of 212 and 168 kJ/mol. For FDH, we have studied the reaction mechanism. Our results indicate that the substrate formate does not coordinate directly to Mo ion when it enters the oxidised active site of FDH, but instead resides in the second coordination sphere. The sulfido ligand abstracts a hydride from substrate, giving a Mo(IV)–SH state. Finally, the CO ₂ will be released when the active site is oxidised by two electrons.			
Key words Metalloenzymes, reaction mechanism, QM/MM, DFT, DMRG, CAS-srDFT, big-QM, QTCP			
Classification system and/or index terms (if any)			
Supplementary bibliographical information		Language English	
ISSN and key title		ISBN 978-91-7422-585-3 (pdf) 978-91-7422-584-6 (print)	
Recipient's notes	Number of pages 116		Price
	Security classification		

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature Geng Dong Date 2018-04-18

Reaction mechanism of metalloenzymes studied by theoretical methods

Geng Dong



LUND
UNIVERSITY

Funding information:

This thesis work was financially supported by China Scholarship Council, the Swedish Research Council and COST through Action CM1305 (ECOSTBio).

Faculty of Science
Division of Theoretical Chemistry

ISBN
978-91-7422-585-3 (pdf)
978-91-7422-584-6 (print)

Printed in Sweden by Media-Tryck, Lund University
Lund 2018



MADE IN SWEDEN 



Media-Tryck is an environmentally certified and ISO 14001 certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

For Xiaoqiong

Contents

List of Publications.....	viii
Popular Science Summary.....	xi
1 Introduction.....	1
1.1 [NiFe] Hydrogenase.....	1
1.2 Dimethyl Sulfoxide Reductase.....	3
1.3 Formate Dehydrogenase.....	4
2 Theory.....	6
2.1 Quantum Mechanics.....	6
2.1.1 The Born–Oppenheimer Approximation.....	7
2.1.2 Spin and the Pauli Exclusion Principle.....	7
2.1.3 Slater Determinants.....	8
2.1.4 Hartree–Fock Method.....	8
2.1.5 Basis Sets.....	10
2.1.6 Post-HF Methods.....	11
2.1.7 Density Functional Theory.....	15
2.1.8 Multiconfigurational Short-Range DFT Method.....	17
2.2 Molecular Mechanics Methods.....	18
2.3 QM/MM Method.....	20
2.4 Big-QM Approach.....	21
2.5 QTCP Method.....	22
3 Summary of the Articles.....	24
3.1 Paper I.....	24
3.2 Paper II.....	25
3.3 Paper III.....	26
3.4 Paper IV.....	27
3.5 Paper V.....	28
3.6 Paper VI.....	29
4 Conclusions and Outlook.....	30
Reference.....	31
Acknowledgements.....	34

List of Publications

This thesis is based on the following papers, which are found at the end of the thesis. They are referred to by Roman numerals.

- I. Protonation states of intermediates in the reaction mechanism of [NiFe] hydrogenase studied by computational methods.
Dong, G. & Ryde, U. *Journal of Biological Inorganic Chemistry*, 2016, 21, 383–394.
- II. H₂ binding to the active site of [NiFe] hydrogenase studied by multiconfigurational and coupled-cluster methods.
Dong, G., Phung, Q. M., Hallaert, S. D., Pierloot, K. & Ryde, U. *Physical Chemistry Chemical Physics*, 2017, 19, 10590–10601.
- III. Exploration of H₂ binding to the [NiFe] hydrogenase active site with multiconfigurational density functional theory.
Dong, G., Ryde, U., Jensen, H. J. Aa. & Hedegård, E. D. *Physical Chemistry Chemical Physics*, 2018, 20, 794–801.
- IV. The full reaction mechanism of [NiFe] hydrogenase studied by computational methods.
Dong, G., Phung, Q. M., Pierloot, K. & Ryde, U.
Manuscript
- V. Effect of the protein ligand in DMSO reductase studied by computational methods.
Dong, G. & Ryde, U. *Journal of Inorganic Biochemistry*, 2017, 171, 45–51.
- VI. Reaction mechanism of formate dehydrogenase studied by computational methods.
Dong, G. & Ryde, U.
Manuscript

List of papers not included in this thesis

1. U. Ryde, G. Dong, J. Li, M. Feldt, R. A. Mata (2016) "Computational studies of molybdenum and tungsten enzymes" In Molybdenum and Tungsten Enzymes, Eds. R. Hille, M. Kirk, C. Schulzke, RSC Metallobiology series no. 7; pp. 275–321. *Book Chapter*.
2. M. Misini Ignjatović, O. Caldararu, G. Dong, C. Muñoz-Gutierrez, F. Adasme-Carreño, U. Ryde, (2016) "Binding-affinity predictions of HSP90 in the D3R Grand Challenge 2015 with docking, MM/GBSA, QM/MM, and free-energy simulations", *J. Comp.-Aided Mol. Design*, 30, 707–730
3. G. Dong, U. Ryde (2016) "O₂ activation in salicylate 1,2-dioxygenase: A QM/MM study reveals the role of His162", *Inorg. Chem.*, 55, 11727–11735
4. G. Dong, L. Cao, U. Ryde (2018) "Insight into the reaction mechanism of lipoyl synthase: A QM/MM study, *J. Biol. Inorg. Chem*, 23, 221–229
5. G. Dong, U. Ryde (2018) "A broken-symmetry DFT study of spin states of noncubane [4Fe–4S] clusters", manuscript

List of article contributions

- I. I performed most of the calculations. I participated in discussion and commented on the manuscript.
- II. I performed most of the calculations. I participated in the data analysis and wrote the first version of the manuscript.
- III. I performed most of the calculations. I participated in the data analysis and wrote the first version of the manuscript.
- IV. I performed most of the calculations. I participated in the data analysis and wrote the first version of the manuscript.
- V. I performed most of the calculations. I participated in the data analysis and wrote the first version of the manuscript.
- VI. I performed most of the calculations. I participated in the data analysis and wrote the first version of the manuscript.

Popular Science Summary

Metalloenzymes are proteins that contain one or more metal ions bound to protein. They constitute about one-third of all enzymes known so far and they often perform hard chemical reactions involving small substrates, like H_2 and N_2 . In our research, we focus on those metal ions that are located in the active sites and perform redox reactions, i.e. involving electron transfer.

Why enzymes? As we know, human benefit from enzymes. For example, more than 700 types of enzymes exist in our body, and O_2 , which is necessary to people, is generated by enzymes. However, the enzymes are complicated and very difficult to understand. In this thesis, theoretical methods were used to investigate enzymes.

Why theoretical methods? Enzymatic reactions are generally fast, so the details of the reaction are hard to study experimentally. Theoretical methods, using computers and software, have become more and more important in the study of enzyme catalytic mechanisms. With theoretical methods, we can construct models to mimic the reaction, so that we can understand the reaction in atomistic details, e.g. electron and proton transfer, bond cleavage and formation, etc. These findings from theoretical studies can then be used in the experimental studies. Here, we study three metalloenzymes, viz. [NiFe] hydrogenase, dimethyl sulfoxide reductase (DMSOR) and formate dehydrogenase.

[NiFe] hydrogenases catalyse the reversible formation of hydrogen molecules from protons and electrons. This very simple reaction has attracted much interest because H_2 may be used as clean and renewable energy carrier. In DMSOR, the reduced enzyme reacts with dimethyl sulfoxide (DMSO) to generate dimethyl sulfide (DMS). This enzyme is interesting because the molybdenum (Mo) is the only known second-row transition metal that employed by proteins, and Mo enzymes exist in almost all organisms and they are involved in the metabolism of many biological systems. Finally, the formate dehydrogenases (FDHs) can react with formate to generate carbon dioxide reversibly. This reaction is a key part of biological transformations of carbon dioxide (CO_2) in the global carbon cycle.

These enzymes are very interesting and play important roles in nature. However, the reaction mechanisms are still not fully understood. In this thesis, we explored the details of the reaction mechanism for the three enzymes with theoretical methods.

1 Introduction

Metalloenzymes exist in all forms of life in nature and play essential roles in the function of all organisms. In this doctoral thesis, three metalloenzymes were studied, viz. [NiFe] hydrogenase (H₂ase), dimethyl sulfoxide reductase (DMSOR) and formate dehydrogenase (FDH), using theoretical methods, namely quantum mechanics (QM), combined quantum mechanical and molecular mechanics (QM/MM), big-QM, as well as QM/MM thermodynamic cycle perturbation (QTCP). The background of the three enzymes will be briefly discussed in this section.

1.1 [NiFe] Hydrogenase

Hydrogenases are metalloenzymes that catalyse the reversible conversion of protons and electrons to H₂ molecules. In nature, three types of hydrogenases are found, categorised according to the metals in their active sites, viz. [Fe], [FeFe] and [NiFe] hydrogenases. Here, we focus on the [NiFe] hydrogenases. Standard [NiFe] hydrogenases are composed of two subunits: the large and small subunits.¹ The [NiFe] active site is located in the large subunit. As shown in Figure 1, the iron ion is coordinated by one carbon monoxide and two cyanide molecules. In addition, two thiolates from Cys84 and 549 bridge the two metals (the residues are numbered according to the enzyme from *Desulfovibrio vulgaris* Miyazaki F).² The nickel ion has two additional cysteine ligands (Cys81 and 546) that are terminally coordinated. The small subunit harbours three FeS clusters in an electron transfer chain, viz. the proximal [4Fe4S], medial [3Fe4S] and distal [4Fe4S] clusters.

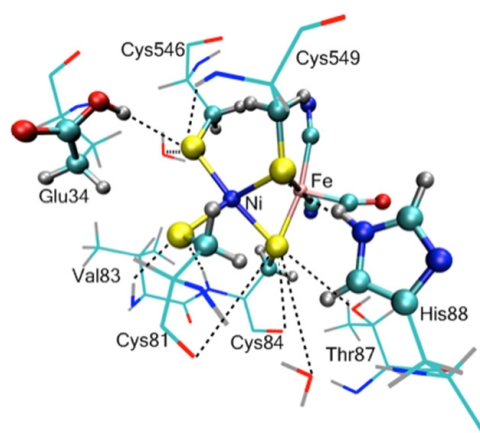


Figure 1. The active site of [NiFe] hydrogenase.

In the catalytic cycle of [NiFe] hydrogenase, a number of intermediates have been identified experimentally.³ The reaction starts from an electron paramagnetic resonance (EPR)-silent state, called Ni-SI_a state, in which the Ni ion is in the +II oxidation state, without any extra ligands. The Fe ion is supposed to remain in the low-spin +II throughout the reaction. Then, one electron and one proton are added to the active site to generate Ni-C state. In the Ni-C state, a hydride ion bridges two metals,^{4,5} and the Ni ion is oxidised to +III. Next, another H⁺/e⁻ pair is added, resulting in the Ni-R state. In the Ni-R state, the hydride and proton react to produce a H₂ molecule, but it still binds in the active site. Finally, the H₂ molecule is released, regenerating the Ni-SI_a state to start a new reaction cycle. Recently, a species namely Ni-L state, which is generated from Ni-C state, was suggested to be an intermediate that may be involved in the catalytic cycle.^{6,7} Based on these studies, a tentative reaction mechanism has been suggested, as shown in Figure 2.

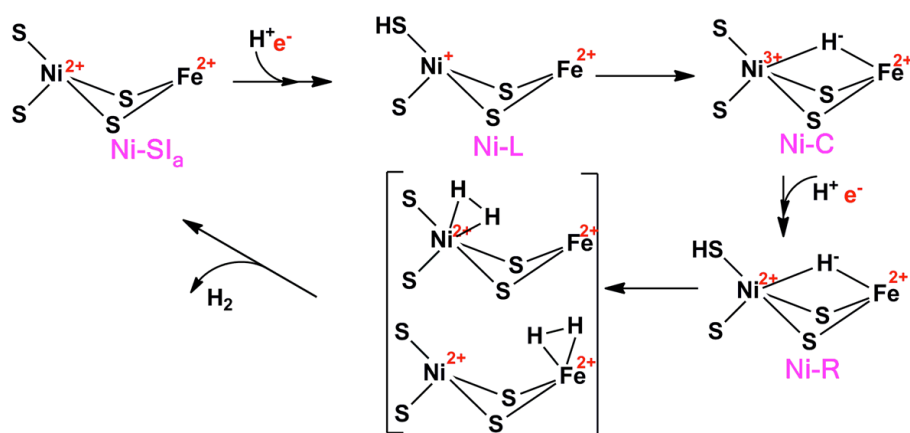


Figure 2. The putative reaction mechanism of [NiFe] hydrogenases.

In the catalytic cycle, two protons and two electrons are required to generate one hydrogen molecule. However, the complete proton transfer mechanism of [NiFe] hydrogenase has not been elucidated. Mutation studies showed that the protons are transferred from the terminal cysteine at the active site to a nearby glutamate residue (Glu34 in Figure 1).⁸ Also, a recent crystallographic study indicated that in an almost (96%) pure Ni-R state, the Cys546 is protonated.⁴

As we can see from Figure 2, the binding site of hydrogen molecule in active site of [NiFe] hydrogenase is unclear. Experimentally, carbon monoxide, which is a competitive inhibitor of H₂, binds to Ni.⁹ Xenon-binding experiment showed a binding path that also ends at Ni ion.^{10,11} In contrast, the Fe ion is suggested as binding site of H₂ from an organometallic perspective.¹² Likewise, theoretical studies have given varying results when different methods and models were used (see Table 1 in paper II). From previous theoretical studies, we can conclude that the energies are very sensitive to the size of the QM

region and the DFT methods. Therefore, advanced methods are required to investigate the H₂ binding mode.

Finally, the details of the hydrogen evolution reaction did not coincide in recent published theoretical studies.^{13,14} The main difference is *whether an extra intermediate exists between the Ni-S_a and Ni-R states, or not*. In addition, the Ni-L state was found by experimental studies in dark environment.^{6,7} However, it is still unclear *Whether it is involved in the catalytic cycle*.

In this thesis, we have investigated the protonation states of the four cysteines in the active site (paper I), the H₂ binding site (papers II and III), and the full reaction mechanism (paper IV) with our theoretical methods.

1.2 Dimethyl Sulfoxide Reductase

Dimethyl sulfoxide reductases (DMSOR) are enzymes that reduce DMSO to DMS by abstracting the oxygen atom of the substrate to a Mo ion (shown in Figure 3).^{15,16} In this process, two electrons are transferred from the Mo ion to the oxygen atom, resulting in the change of oxidation state from +IV to +VI.

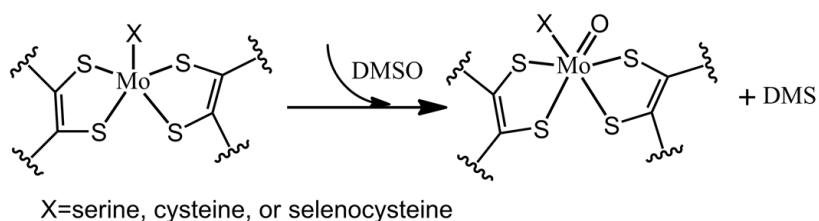


Figure 3. The overall reaction of DMSOR.

The active site of reduced DMSOR contains two molybdopterin (MPT) cofactors bound to the Mo(IV) ion in a nearly planar fashion (Figure 4),^{15,16} and one deprotonated side-chain O, S or Se atom from serine, cysteine or selenocysteine at the apical position. The reaction mechanism of DMSOR with serine binding to Mo ion has been thoroughly studied.¹⁷⁻²⁸ The product DMS is generated by a two-step reaction: 1) the substrate DMSO binds to Mo(IV) ion; 2) two electron transfer from the Mo(IV) ion to the substrate as the S–O bond is cleaved. All studies indicated that the second reaction is the rate-determining step with a barrier of 38–80 kJ/mol.^{18-23,28,29} Studies from our group have shown that calculated barrier strongly depends on the theoretical method and that a proper account of dispersion and solvation effects is needed, together with large basis sets and accurate density functional theory (DFT) methods.^{17,25}

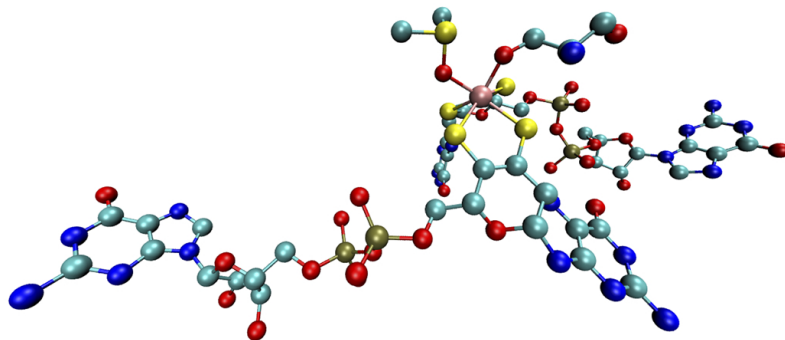


Figure 4. The active site of DMSO reductase, showing that the Mo ion coordinates to two MPT cofactors, DMSO and a Ser residue (PDB ID: 4DMR).

However, how the reaction mechanism and rate change when DMSO reacts with the enzymes with different protein-derived ligands (serine, cysteine and selenocysteine) is still unclear. In paper V, small models were used to systematically investigate the reaction mechanism with various protein-derived ligands.

1.3 Formate Dehydrogenase

Biological transformations of carbon dioxide (CO_2) are key processes in global carbon cycle and have attracted great attention, because they may be used to combat the greenhouse effect. In nature, the reversible conversion of CO_2 to formate (HCOO^-) is catalysed by formate dehydrogenases (FDHs).^{15,30-33} The FDHs can be divided into two classes, metal-dependent and metal-independent. The metal-dependent FDHs can be further classified into molybdenum- and tungsten-FDHs. Here, we focus on Mo-FDHs. Similar to DMSOR, the active sites of FDHs in the oxidised (Mo^{VI}) state contain two molybdopterin cofactors (MPT, Figure 4), one protein-derived ligand (cysteine or selenocysteine) and a sulfido group, which coordinate to Mo ion in a distorted hexa-coordinated trigonal prismatic geometry.³⁴⁻³⁸

As the FDHs play important roles in the global carbon cycle, the reaction mechanism has been extensively studied by both experimental and computational methods.^{15,31,34,39} However, the conclusions from these studies did not coincide. As shown in Figure 5, five mechanisms have been suggested (see the detailed discussion in paper VI). The main issue is *whether the substrate binds to Mo ion and whether the protein-derived ligand (cysteine or selenocysteine) dissociates from Mo during the reaction mechanism.*

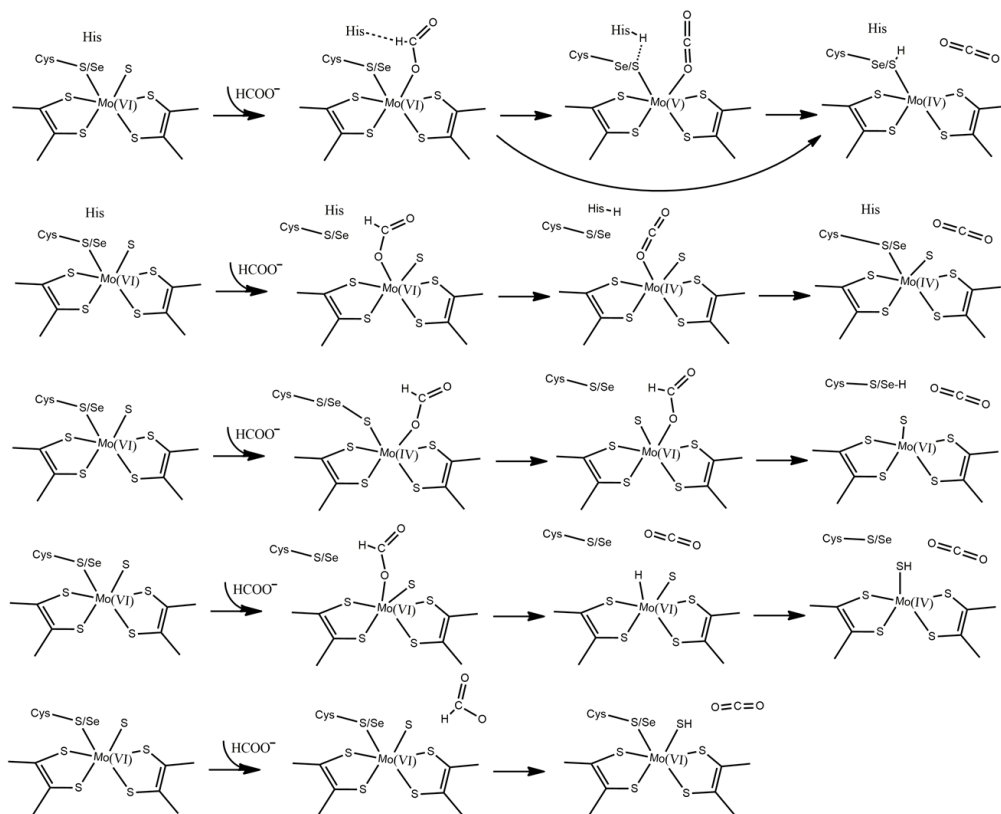


Figure 5. Suggested reaction mechanisms of FDHs.

In paper VI, we examined all these reaction mechanisms and discussed these two questions. Finally, a reasonable mechanism was proposed.

2 Theory

In this chapter, I will give a basic introduction to the theoretical methods used in the thesis, viz. QM-cluster, QM/MM, big-QM and the QM/MM thermodynamic cycle perturbation (QTCP) methods. For the QM-cluster approach, density functional theory (DFT), density matrix renormalisation group (DMRG), coupled cluster (CC) and short-range DFT (srDFT) calculations were employed in our papers. All these methods will be briefly introduced in order to give an overview of the underlying concepts. All equations are expressed in atomic units, i.e. \hbar (Planck constant) = m_e (mass of one electron) = e (charge of electron) = k_e (Coulomb's constant, $1/4\pi\epsilon_0$) = 1.

2.1 Quantum Mechanics

In quantum mechanical (QM) methods, a system is described by the Schrödinger equation, which in the time-independent form is

$$\hat{H}\Psi = E\Psi \quad (1)$$

where Ψ is the wave function, which is a function of all electron and nuclear positions in the system. Its squared absolute value, $|\Psi|^2$, corresponds to the probability distribution function for particles to be found in a specified volume element. \hat{H} is the Hamiltonian operator, and E is the total energy of the system. The Hamiltonian is defined by

$$\begin{aligned} \hat{H} &= \hat{T}_e + \hat{T}_N + \hat{V}_{Ne} + \hat{V}_{ee} + \hat{V}_{NN} \\ &= -\frac{1}{2} \sum_{i=1}^n \nabla_i^2 - \frac{1}{2} \sum_{A=1}^N \frac{1}{M_A} \nabla_A^2 - \sum_{i=1}^n \sum_{A=1}^N \frac{Z_A}{|r_i - R_A|} + \sum_i \sum_{j>i} \frac{1}{|r_i - r_j|} \\ &\quad + \sum_A \sum_{A>B} \frac{Z_A Z_B}{|R_A - R_B|} \end{aligned} \quad (2)$$

where A and B denote nuclei, whereas i and j denote electrons. n is the number of electrons, whereas N is the number of nuclei. M_A is the mass and Z_A is the charge of the nucleus. ∇_i^2 (the Laplace operator) is the gradient with respect to coordinates of particle i . r_i and R_A are the positions of electron i and nucleus A , respectively. $|r_i - R_A|$, $|r_i - r_j|$ and $|R_A - R_B|$ are the distances between electron and nucleus, electron and electron, nucleus and nucleus, respectively.

The first and second terms of Eq. 2 are the kinetic energies of electrons and nuclei, respectively. The remaining three terms are potential energies, and they are, from left to right, the electron–nucleus attraction, the electron–electron repulsion and the nucleus–nucleus repulsion. Unfortunately, the Schrödinger equation can be analytically solved only for a few simple systems, e.g. systems with one nucleus and a single electron — for all large systems, only approximate numerical solutions can be obtained.

2.1.1 The Born–Oppenheimer Approximation

Solving Schrödinger equation can be simplified with the Born–Oppenheimer approximation, which was introduced by Born and Oppenheimer in 1927.⁴⁰ In this approximation, the key idea is the separation of electronic and nuclear motions. The mass of the nucleus is three to five orders of magnitude larger than the mass of an electron, resulting in a much faster movement of electrons compared to nuclei. Therefore, the movement of nuclei and the coupling between the nuclei and electronic motion can be neglected (the second term of Eq. 2), while the nuclear repulsion can be considered to be constant (the last term of Eq. 2). Thus, we obtain

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^n \nabla_i^2 - \sum_{i=1}^n \sum_{A=1}^N \frac{Z_A}{|r_i - R_A|} + \sum_i \sum_{j>i} \frac{1}{|r_i - r_j|} + \sum_A \sum_{A>B} \frac{Z_A Z_B}{|R_A - R_B|} \quad (3)$$

which is the electronic Schrödinger equation. Note that the nuclei are only present as constant parameters (charges and positions).

2.1.2 Spin and the Pauli Exclusion Principle

In quantum mechanics, spin is an intrinsic property of particles (without any classical analogue), and it behaves like an angular momentum. Particles with half-integer spins, such as $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, are known as fermions, and those with integer spins are known as bosons. Electrons are fermions with a half-integer spin ($s = 1/2$). Two spin states can be adopted for one electron: the state with $m_s = +1/2$ is called $|\alpha\rangle = \alpha(\omega)$, and the state with $m_s = -1/2$ is called $|\beta\rangle = \beta(\omega)$, where ω is the spin coordinates.

The Pauli exclusion principle states that two or more identical fermions cannot occupy the same quantum state simultaneously. For a many-particle wave function the Pauli exclusion principle corresponds to that the wave

function is antisymmetric (for fermions) with respect to interchange of coordinates any two fermions x_i and x_j . This principle imposes an additional constraint on the wave function.

2.1.3 Slater Determinants

One possible wave function ansatz for N -electron system problem is to construct the wave function as a Slater determinant

$$\begin{aligned} & \Psi^{SD}(x_1, x_2, x_3, \dots, x_N) \\ &= \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) & \dots & \chi_{N-1}(x_1) & \chi_N(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \dots & \chi_{N-1}(x_2) & \chi_N(x_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \chi_1(x_{N-1}) & \chi_2(x_{N-1}) & \dots & \chi_{N-1}(x_{N-1}) & \chi_N(x_{N-1}) \\ \chi_1(x_N) & \chi_2(x_N) & \dots & \chi_{N-1}(x_N) & \chi_N(x_N) \end{vmatrix} \end{aligned} \quad (4)$$

Thereby, the total N -electron wave function obeys the Pauli exclusion principle and displays the correct antisymmetry, because the determinant is zero if any two rows or columns are same and changes the sign upon exchanging two rows. In the determinant, χ is a one-electron wave function, also known as a spin-orbital, which consist of a spatial orbital $\psi(r)$, multiplied by a spin function

$$\chi = \psi(r) \cdot \begin{cases} \alpha(\omega) \\ \beta(\omega) \end{cases} \quad (5)$$

The MOs are orthonormal, $\langle \chi_i | \chi_j \rangle = \delta_{ij}$.

2.1.4 Hartree–Fock Method

The Hartree–Fock (HF) method is one of the most common approximations to solve the Schrödinger equation.^{13,41-43} Here, a single-determinant in Eq. 4 is used in the Schrödinger equation. Thus, each electron moves in the average potential of all the other electrons, which means that N -electron problem is converted to a set of one-electron problems, i.e. the electron–electron repulsion is calculated in an average way.

Here, I will focus on the restricted Hartree–Fock theory for closed-shell molecules, in which the spatial function $\psi(r)$ in Eq. 5 is same for the α and β electrons in the same orbital. Thus, the many-particle electronic Hamiltonian is replaced by the Fock operator \hat{f} :

$$\hat{f} = \hat{h}_i + \hat{V}_i = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^N \frac{Z_A}{|r_i - R_A|} + V_i^{HF} \quad (6)$$

$$V_i^{HF} = \sum_{j=1}^{N/2} (2\hat{J}_j - \hat{K}_j) \quad (7)$$

In Eq. 6, the first and second terms of right hand side are the electron kinetic energy and the nucleus–electron attraction, respectively. V_i^{HF} is the Hartree–Fock potential, i.e. the mean field from all electrons besides electron i . \hat{J}_j is the Coulomb operator, defining the electron–electron repulsion between electron i and j . \hat{J}_j is multiplied by 2 to account for the presence of two electrons in each orbital. \hat{K}_j is the exchange operator, which represents the energy associated with exchanging two electrons.

\hat{J} and \hat{K} operate on a wave function χ_i , which describes the interaction of electron 1 (one-electron Coulomb and exchange operators),

$$\hat{J}_j(1)\chi_i(1) = \left[\int \chi_j^*(2) \frac{1}{r_{12}} \chi_j(2) d\tau_2 \right] \chi_i(1) \quad (8)$$

$$\hat{K}_j(1)\chi_i(1) = \left[\int \chi_j^*(2) \frac{1}{r_{12}} \chi_i(2) d\tau_2 \right] \chi_j(1) \quad (9)$$

In Eq. 8, the quantity $\chi_j^*(2) \frac{1}{r_{12}} \chi_j(2) d\tau_2$ presents the potential energy of one electron at r_1 due to the charge density at r_2 , where r_{12} is the distance between r_1 and r_2 . Evaluation of the integral gives the total potential energy at r_1 due to the overall (or average) charge density produced by electron 2 in orbital j . The solution of the Hartree–Fock equation produces a spin orbital χ_i that is determined by the average potential energy (or Coulomb field) of all the other electrons. In Eq. 9, $\chi_j^*(2) \frac{1}{r_{12}} \chi_i(2) d\tau_2$ represents the potential energy at r_1 due to the overlap charge distribution at r_2 associated with orbitals i and j . The integral gives the potential energy due to the total overlap charge density associated with electron 2. However, there is no classical analogue to the exchange energy and it is introduced as the result of the antisymmetry of wave function.

The orbitals are eigenfunctions of the Fock operator, and the eigenvalue ε_i is the orbital energy.

$$\hat{f}_i \chi_i = \varepsilon_i \chi_i \quad (10)$$

The total electron energy is not the sum of orbital energies, because the Fock operator contains terms describing the repulsion to all other electrons (\hat{J} and \hat{K}), and the sum over MO energies therefore counts the electron–electron repulsion twice.⁴⁴ Instead, we can write

$$E = \sum_{i=1}^n \varepsilon_i - \frac{1}{2} \sum_j^n V_i^{HF} \quad (11)$$

2.1.5 Basis Sets

To solve the set of equations in Eq. 10, we need a description of the orbitals (χ). For molecules, the molecular orbitals (MOs), χ , can be constructed by a set of atomic orbitals (AOs, φ). We denote the AOs as “basis set”, which is known as the linear combination of AOs method, LCAO and usually refers to the set of (non-orthogonal) one-particle functions (which are known beforehand) to build MOs. Typically, basis functions are centred on the atoms. However, it should be noted that basis functions are usually not true atomic orbitals. In general, larger basis sets lead to a more accurate result, but also an increased computational cost.

The employed AOs can either be Slater-type orbitals (STO)⁴⁵ or Gaussian-type orbitals (GTO).⁴⁶ The STOs are defined by

$$\varphi_{nlm}^{STO}(r, \theta, \phi) = Nr^{n-1}e^{-\zeta r}Y_l^m(\theta, \phi) \quad (12)$$

where n , l , and m are the electron quantum numbers. N is a normalisation constant, r is the distance between the electron and atomic nucleus, ζ is a constant, which is related to the effective charge of nucleus (the nuclear charge is partly shielded by electrons) and controls the width of orbitals (large ζ gives tight function and small ζ gives diffuse functions). Y_l^m is the spherical harmonics functions.

The Gaussian-type orbitals are most often used, because they are more efficient in numerical calculations. They have the form

$$\varphi_{nlm}^{GTO} = Nr^{n-1}e^{-\zeta r^2}Y_l^m(\theta, \phi) \quad (13)$$

STOs are more accurate, but the involved integrals are more complicated. To mimic the STOs, a strategy is to use a linear combination of several GTOs, called contracted GTOs (CGTO). The STO-3G is a well-known minimal basis set, which contracts three GTOs to mimic the STOs. CGTO might give a good approximation for an atomic orbital, but it lacks the flexibility to expand or shrink in the presence of other atoms in a molecule.

Therefore, they cannot give highly accurate results. To improve this approximation, more than one CGTO is often used for each electron. For example, doubling the number of GTO per electron is termed a double- ζ basis set. In practice, we often use a split-valence basis set, i.e. one CGTO is used for core electrons, but two for the valence electrons. Furthermore, so-called polarisation and diffuse functions can be added to improve the accuracy. In general, the polarisation functions are represented by GTOs of angular momentum $l+1$. Diffuse functions are required to use for the description of anions and the systems with electron distributions that extend further from the nuclei. Diffuse functions have small ζ exponents to hold the electrons far from the nucleus.

In this thesis, Karlsruhe,⁴⁷ Pople,⁴⁸ and atomic natural orbital with relativistic and core correlation (ANO-RCC)-type⁴⁹ basis sets are used. Among the Karlsruhe basis sets, def2-SV(P) is a valence double-zeta basis set with polarisation functions on heavy atoms; def2-TZVP is valence triple-zeta polarisation basis set; and def2-QZVPD is valence quadruple-zeta polarisation basis set with diffuse functions. The Pople-type basis sets are labelled as X - YZG (e.g. 6-31G), where X is the number of GTOs linearly combined to construct CGTOs for the core electrons, Y and Z indicate that the valence electrons are described by two basis function (CGTOs), i.e. a valence double-zeta basis set (likewise, X - $YZWG$ means a triple-zeta basis set). For ANO-RCC basis set, ANO-RCC-VZDP, ANO-RCC-VZTP, and ANO-RCC-VZQP are valence double-, triple- and quadruple-zeta basis sets, respectively.

2.1.6 Post-HF Methods

The HF method adopts a mean-field approximation, i.e. the electron–electron repulsion is not described rigorously. However, the accuracy of calculation with this method is not good enough for many cases. Therefore, several methods use the HF wave function as the starting point and then add a more elaborate description of electron–electron interactions (electron correlation).

2.1.6.1 Møller–Plesset Perturbation Theory

Møller–Plesset perturbation theory is a widely used method in quantum chemistry calculations.⁵⁰ Adopting perturbation theory, we can write the Hamiltonian as

$$\hat{H} = \hat{H}_0 + \lambda\hat{V} \tag{14}$$

In Møller–Plesset perturbation theory, we set \hat{H}_0 as the Fock operator and the perturbation to \hat{V} , and λ is the perturbation strength. In HF theory, the sum of Fock operators counts the electron–electron repulsion twice, thus the

perturbation becomes the exact \hat{V}_{ee} operator minus twice the $\langle \hat{V}_{ee} \rangle$ operator, i.e. $\hat{V} = \hat{V}_{ee} - 2\langle \hat{V}_{ee} \rangle$.

The zero-order equation is simply the Schrödinger equation for the unperturbed system and the corresponding energy is

$$E^{(0)} = \langle \Psi^{(0)} | \hat{H}_0 | \Psi^{(0)} \rangle \equiv E(\text{MP0}) \quad (15)$$

This is just a sum of MO energies. The first-order correction to the energy is

$$E^{(1)} = \langle \Psi^{(0)} | \hat{V} | \Psi^{(0)} \rangle \equiv E(\text{MP1}) \quad (16)$$

This correction is for the overcounting of the electron–electron repulsion at zeroth order. Thus, the sum of $E(\text{MP0})$ and $E(\text{MP1})$ is the HF energy. The second-order (MP2) energy, which is most used approximation, can be expressed as

$$E^{(2)} = \sum_t \frac{|\langle \Psi^{(0)} | \hat{V} | \Psi_t \rangle|^2}{E^{(0)} - E_t} \quad (17)$$

2.1.6.2 Configurational Interaction Methods

Another post-HF method is configurational interaction (CI).^{51,52} In order to account for electron correlation, a variational wave function is constructed by linear combination of configuration state function (CSFs) built from spin orbitals,

$$\Psi = \sum_{i=0} C_i \Psi_i = C_0 \Psi_0 + C_1 \Psi_1 + C_2 \Psi_2 + \dots \quad (18)$$

The first term Ψ_0 is normally the HF determinant and the other CSFs represent the determinants with some electrons excited to virtual orbitals: If only one-electron excitations are included (swapping one occupied spin orbital with virtual orbital in the determinant), it is called CIS; and if only two-electron excitation is allowed, it is called CID. The most often-used method is CISD which is limited to single and double excitations. The method with all possible determinants considered (with a certain basis set) is called full CI (FCI), which represents the exact solution within that basis set. In quantum chemistry calculation, the energy is minimised by varying the coefficients in Eq. 18.

2.1.6.3 Coupled Cluster Methods

The coupled cluster (CC) method is an accurate method, but with very high computational cost.⁵³ In particular, the singles and doubles with perturbatively treated triples (CCSD(T)) has become the current gold-standard of quantum chemistry. The wave function of the CC theory is written in terms of exponential functions

$$|\Psi\rangle = e^{\hat{T}} |\Psi_0\rangle \quad (19)$$

where Ψ_0 is a Slater determinant usually constructed from the HF wave function. \hat{T} is an excitation operator that is a linear combination of excited Slater determinants. The \hat{T} is expressed as

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots \quad (20)$$

where \hat{T}_1 is the operator of all single excitations, \hat{T}_2 is the operator of all double excitations, and so forth. The exponential operator $e^{\hat{T}}$ can be written in form of Taylor series

$$e^{\hat{T}} = 1 + \hat{T} + \frac{\hat{T}^2}{2!} + \dots = 1 + \hat{T}_1 + \hat{T}_2 + \frac{\hat{T}_1^2}{2} + \hat{T}_1\hat{T}_2 + \frac{\hat{T}_2^2}{2} + \dots \quad (21)$$

In practice, the expansion of \hat{T} into individual excitation operators is usually terminated at the second level of excitation. Coupled cluster methods usually recover more correlation energy than CI methods with the same maximum excitations, due to the non-linear nature of exponential function.

2.1.6.4 Complete Active Space Methods

All the mentioned post-HF methods above assume that the HF Slater determinant is a qualitatively correct reference wave function and thus the correlation is small. However, for e.g. near-degenerate ground states and bond-breaking reactions, the HF approximation becomes problematic. These cases are said to introduce static correlation. Static correlation means in other words that more than one determinant has a large weight in the total wave function. The post-HF methods described up to here, e.g. CCSD(T), treat the dynamic correlation (which is described by a (large) number of determinants with small weight) well. However, these methods usually fail for static correlation. The multi-configurational self-consistent field (MCSCF) methods can be considered as a CI where not only the coefficients of Eq. 18 are optimised, but also the MOs used for constructing the determinants are optimised.⁵⁴ Thus, the Ψ_i in Eq. 18 are now considered as $\Psi_i = |\chi_1\chi_2\chi_3 \dots \chi_N|$, where

$$\chi_i = \sum_{\mu} C_{i\mu} \Phi_{\mu} \quad (22)$$

The $C_{i\mu}$ are the MO coefficients. Both the CI and MO coefficients, C_i and $C_{i\mu}$, are optimised in MCSCF. With the selected configurations, the MCSCF can generate a qualitatively correct wave function, i.e. recovering the static part of the correlation.

One of most commonly used MCSCF methods is called the complete active space self-consistent field (CASSCF).⁵⁵ The selection of configurations is done by partitioning the MOs into active and inactive spaces. In the active space, a full CI is performed, i.e. all possible excitations are considered. Using the reference wave function from CASSCF calculation, complete active space perturbation theory (CASPT2) can be employed to obtain the dynamic correlation in a similar manner as MP2 is applied on the HF wave function.⁵⁶

In CASSCF calculation, only a small number of orbitals and electrons can be handled (currently up to 18 electrons in 18 orbitals) in the active space. In 1992, the density matrix renormalisation group (DMRG) theory was introduced.⁵⁷ This method can solve the CASSCF calculations with a significant larger active space (up to about 50 electrons in 50 orbitals).

DMRG arranges the orbitals (“sites”) in the CAS linearly. Each orbital has a physical degree of freedom $|\sigma_j\rangle$. We can write the wave function as

$$|\Psi\rangle = \sum_{\sigma_1 \dots \sigma_L} C_{\sigma_1 \dots \sigma_L} |\sigma_1 \dots \sigma_L\rangle \quad (23)$$

Similar to Eq. 18, but the coefficients in Eq. 23 are written as a coefficient tensor. The tensor $C_{\sigma_1 \dots \sigma_L}$ can be rewritten in matrix product states (MPSs) using a series of singular-value decompositions (SVD) as

$$|\Psi\rangle = \sum_{\sigma_1 \dots \sigma_L} M_{1,a_1}^{\sigma_1} M_{a_1,a_2}^{\sigma_2} \dots M_{a_{L-1},1}^{\sigma_L} |\sigma_1 \dots \sigma_L\rangle \quad (24)$$

The DMRG algorithm optimises the site matrices iteratively. $M_{a_{i-1},a_i}^{\sigma_i}$ is $D \times D$ matrix for each site (rank-3 tensor), where D is also the dimension of the MPS (virtual dimension). The matrix D determines how accurate the DMRG approximates the full CAS. Values of 1000–2000 are usually sufficient but this is can be system-dependent.

2.1.7 Density Functional Theory

The original idea of density functional theory (DFT) is to characterise systems by the electron density $\rho(r)$, which has only 3 variables, viz. the three Cartesian coordinates x , y , and z . The computational effort to solve Schrödinger equation is thus significantly reduced, compared to the wave function methods. However, if all the energy components are expressed as a functional of the electron density, it turns out the DFT methods give poor results.⁴⁴ Nowadays, DFT is based on Kohn-Sham theory, in which the electron kinetic energy is calculated from an auxiliary set of orbitals used to represent the electron density, meaning that a dependence on orbitals is introduced.

2.1.7.1 Kohn–Sham DFT

Kohn and Sham formulated a variant DFT that use the framework of the Hartree–Fock method,⁵⁸ and the Kohn-Sham electronic energy is expressed as

$$E(\rho) = T_e(\rho) + E_{ne}(\rho) + J_{ee}(\rho) + E_{XC}(\rho) \quad (25)$$

where T_e is the kinetic energy of the non-interacting electrons, E_{ne} is the electron–nucleus attraction, J_{ee} is the electron repulsion, and the last term is known as exchange–correlation energy that will be discussed in next section. The one-electron Kohn–Sham operator is

$$\hat{h}_i^{KS} = -\frac{1}{2} \sum_{i=1}^n \nabla_i^2 - \sum_{i=1}^n \sum_{A=1}^N \frac{Z_A}{|r_i - R_A|} + \int \frac{\rho(r')}{|r_i - r'|} dr' + V_{XC} \quad (26)$$

The Kohn–Sham electron density $\rho(r)$ is obtained from the Slater determinant,

$$\rho(r) = \sum_{i=1}^n |\chi_i(r)|^2 \quad (27)$$

and the exchange–correlation potential, V_{XC} , is defined as

$$V_{XC} = \frac{\delta E_{XC}}{\delta \rho} \quad (28)$$

The eigenvalue equation we solve in the Kohn–Sham method is

$$\hat{h}_i^{KS} \chi_i = \varepsilon_i \chi_i \quad (29)$$

where ε_i is the corresponding orbital energy.

In DFT methods, the exchange term \hat{K} in HF theory disappears and the exchange–correlation energy (E_{XC}) is considered by the V_{XC} term.

2.1.7.2 Exchange–Correlation Functionals

The only unknown part of Eq. 25 is the exchange–correlation (XC) functional, which can be split into exchange and correlation parts. There are many approximations to the exchange and correlation potentials $v_X(r)$ and $v_C(r)$, respectively.⁵⁹

A basic approximation is the so-called local-density approximation (LDA), which assumes that the density is slowly varying and the inhomogeneous density of a molecular can be approximated using the homogeneous electron gas. The exchange–correlation energy is

$$E_{XC}^{LDA} = \int \rho(r) \varepsilon_{XC}(\rho) dr = \int \rho(r) (\varepsilon_C(\rho) + \varepsilon_X(\rho)) dr \quad (30)$$

where $\varepsilon_{XC}(\rho)$ is the exchange–correlation energy density (the exchange–correlation energy per particle of a uniform electron gas with density ρ). In LDA, the $\varepsilon_{XC}(\rho)$ is a functional of the density only. The corresponding exchange–correlation potential is

$$v_{XC}^{LDA} = \frac{\delta E_{XC}^{LDA}}{\delta \rho(r)} = \varepsilon_{XC}[\rho(r)] + \rho(r) \frac{\partial \varepsilon_{XC}(\rho)}{\partial \rho} \quad (31)$$

For the homogeneous electron gas, the exchange–energy functional can be expressed exactly as

$$\varepsilon_X = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{\frac{1}{3}} \quad (32)$$

This is no explicit expression for the correlation part, but it can be obtained from quantum Monte Carlo simulations.

Another approximation for the exchange–correlation energy is based on generalised gradient approximation (GGA), according to which the exchange–correlation functional depends on both ρ and $\nabla \rho$ (i.e. the first derivative of ρ , also called the gradient charge density). One commonly used GGA functional is that of Becke,

$$v_{XC}^B = -\beta \rho^{\frac{1}{3}} \frac{x^2}{(1 + 6\beta x \sinh^{-1} x)} \quad (33)$$

where $x = \frac{\nabla\rho}{\rho^{4/3}}$, and $\beta=0.0042$, which is determined based on the best fit to the energies of six noble gas atoms using the sum of the LDA and GGA exchange terms.

At the next level of approximation, the meta-GGAs improve the accuracy by employing also the Laplacian (second derivative) of ρ . In practice, they usually include the kinetic energy density (τ) instead of the Laplacian because it is numerically more stable.

$$\tau(\mathbf{r}) = \sum_{i=1}^{N_{occ}} \frac{1}{2} |\nabla\chi_i(\mathbf{r})|^2 \quad (34)$$

Finally, the hybrid functionals add a fraction of HF exchange to $E_X(\rho)$. They have the general form

$$E_{XC}^{hybrid} = (1 - \alpha)E_{XC}^{DFT} + \alpha E_{XC}^{HF} \quad (35)$$

One of the most used hybrid function is B3LYP, which can be expressed as

$$E_{XC}^{B3LYP} = (1 - a)E_X^{LDA} + \alpha E_X^{HF} + b\Delta E_X^B + (1 - c)E_C^{LDA} + cE_C^{LYP} \quad (36)$$

where $a = 0.20$, $b = 0.72$ and $c = 0.81$.

2.1.8 Multiconfigurational Short-Range DFT Method

In this section, a hybrid method, combining wave function theory (WFT) and density functional theory (DFT), will be introduced. This method relies on the range-separation of the two-electron repulsion operator into long-range and short-range parts⁶⁰

$$\hat{V}_{ee}(1,2) = \hat{V}_{ee}^{lr}(1,2) + \hat{V}_{ee}^{sr}(1,2) \quad (37)$$

Several forms of range-separated operators have been developed. Here, we discuss one type range-separated operator that was used in our calculations, which is based on the error function (erf)^{61,62}

$$\hat{V}_{ee}^{lr}(1,2) = \frac{\text{erf}(\mu|r_1 - r_2|)}{|r_1 - r_2|}; \hat{V}_{ee}^{sr}(1,2) = \frac{1 - \text{erf}(\mu|r_1 - r_2|)}{|r_1 - r_2|} \quad (38)$$

where μ is the range-separation parameter. In the limiting case, a value of $\mu = \infty$ implies that the DFT (short-range) part will vanish, giving a pure wave function method; whereas $\mu = 0$ results in a pure Kohn–Sham DFT method. The effective electronic Hamiltonian used in multiconfigurational short-range DFT (MC-srDFT) is

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^n \nabla_i^2 - \sum_{i=1}^n \sum_{A=1}^N \frac{Z_A}{|r_i - R_A|} + \sum_i \sum_{j<i} \frac{\text{erf}(\mu|r_i - r_j|)}{|r_i - r_j|} + \hat{V}_{HXC}^{sr} \quad (39)$$

where \hat{V}_{HXC}^{sr} is short-range adapted and μ -dependent exchange–correlation (XC) potential, obtained from the DFT theory.

It should be stressed that the special exchange–correlation functionals are a prerequisite for range-separated method. In our calculations, the short-range PBE-based srPBE functional was used.^{63,64} For the long-range part, the wave function was described by CASSCF, so-called CAS-srPBE.

2.2 Molecular Mechanics Methods

QM methods, which solve the Schrödinger equation, can at most handle ~ 1000 atoms, e.g. at TPSS/def2-SV(P) level of theory. However, proteins typically contain tens of thousands of atoms, so for such systems the computational cost of a pure QM calculation would be too high. With the molecular mechanics (MM) method, proteins and other large systems can be simulated, because no attempt is made to solve the Schrödinger equation and electrons are ignored. With MM, molecules are described as a collection of balls, connected by springs, and the system is described by an empirical function, a force field.

For proteins, the potential energy typically contains terms for the distortion of bonds, angles, and dihedrals (torsions), as well as the nonbonded exchange-repulsion, dispersion (van der Waals interaction), and electrostatic interaction energies.

$$U_{\text{tot}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{tors}} + U_{\text{vdW}} + U_{\text{elec}} \quad (40)$$

Stretching a covalent bond is normally assumed to be harmonic. Thus, the bond energies can be expressed as

$$U_{\text{bond}} = \sum k_b (r - r_0)^2 \quad (41)$$

where k_b is the bond force constant and r_0 is reference bond length. Likewise, the angle term, U_{angle} , is also approximately described by harmonic potential,

$$U_{\text{angle}} = \sum k_a (\theta - \theta_0)^2 \quad (42)$$

where k_a is the angle force constant, and θ_0 is the reference angle. The dihedral term, U_{tors} , is associated with the rotation around a bond, for which a periodic function is used:

$$U_{\text{tors}} = \sum \frac{1}{2} V_n [1 + \cos(n\omega + \varphi_n)] \quad (43)$$

where V_n is the force constant, n is the periodicity, φ_n is the phase of the torsion, and ω is the torsional angle. The U_{vdW} term uses a Lennard-Jones potential, which divides the interaction into a short-range repulsive term and long-range attractive term,

$$U_{\text{vdW}} = \sum \frac{C_1}{r^{12}} - \frac{C_2}{r^6} \quad (44)$$

where C_1 and C_2 are coefficients that depend on the atom types. The r^{-12} term describe the short-range interactions, i.e. exchange-repulsion due to the overlap of atoms. The r^{-6} term is the long-range interaction, i.e. the dispersion or van der Waals force. The last term, U_{elec} is the Coulomb potential

$$U_{\text{elec}} = \sum \frac{Q_1 Q_2}{4\pi\epsilon\epsilon_0 r} \quad (45)$$

where Q_1 and Q_2 are the charges of the atoms, ϵ_0 is the permittivity of vacuum, ϵ is relative permittivity (also known as the dielectric constant) of the medium, and r is the distance between the atoms.

With MM method, the total energy of a protein can be calculated in seconds. Therefore, molecular dynamics (MD, based on Newton's second law of motion) or Monte Carlo (MC, based on random sampling) can be run to study thermodynamic ensembles of structures.

In addition, the protein is typically solvated with several thousands of explicit water molecules. Several water models can be used in the simulation, e.g. SPC (simple point charge), TIP3P (transferable intermolecular potential with three points), TIP4P, etc. The TIP3P model was used in the thesis,⁶⁵ in which the potential energy, ϵ_{mn} is determined by intermolecular interaction between the sites, described by Coulomb and Lenard-Jones potentials,

$$\epsilon_{mn} = \sum_i^{on} \sum_j^{m \quad on} \left[\frac{Q_i Q_j e^2}{4\pi\epsilon\epsilon_0 r_{ij}} + \frac{A}{r_{00}^{12}} - \frac{C}{r_{00}^6} \right] \quad (46)$$

where, Q is the charge of hydrogen or oxygen atom, and $Q_O = -2Q_H$. A Lenard-Jones potential is used to describe the oxygen–oxygen interaction between two water molecules. In TIP3P, $r_{OH} = 0.9572 \text{ \AA}$, $\angle HOH = 104.52^\circ$, $A = 5.82 \times 10^5 \text{ kcal \AA}^{12}/\text{mol}$, $C = 595.0 \text{ kcal \AA}^6/\text{mol}$, $q_O = -0.834$, and $q_H = 0.417$. The interactions between water and protein are described by the normal MM nonbonded terms (the fourth and fifth terms in Eq. 40). In addition, counter ions are often added to produce a neutral system. To mimic an infinite system, periodic boundary conditions are employed and long-range electrostatic interactions are often treated by Ewald summation.

2.3 QM/MM Method

As discussed above, the QM methods that calculate molecular electronic structure can give good description of the reaction for small systems, whereas the MM method can handle a large system but does not give a good description of the chemical reactivity because electrons are neglected, and e.g. Eq. 41 shows that the bonds cannot be cleaved. The QM/MM approach takes advantage of the accuracy of QM methods and the speed of MM method:⁶⁶ For a small but interesting part where e.g. a chemical reaction occurs, QM calculation is used, whereas the remainder of the protein and the surrounding solvent are described by MM method.⁶⁷ Therefore, the QM/MM methods are widely used for enzymatic reactions.

The total energy of the combined QM/MM methods can be expressed as

$$E_{\text{tot}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM-MM}} \quad (47)$$

where E_{QM} is the energy of the QM region, E_{MM} is the energy of the MM region calculated by an MM force field, and $E_{\text{QM-MM}}$ is the interaction between the QM and MM parts. The $E_{\text{QM-MM}}$ typically consists of three types of interactions: electrostatic interactions ($E_{\text{ES}}^{\text{QM-MM}}$), van der Waals interactions ($E_{\text{vdW}}^{\text{QM-MM}}$), and MM-bonded interactions ($E_{\text{bonded}}^{\text{QM-MM}}$).

$$E_{\text{QM-MM}} = E_{\text{ES}}^{\text{QM-MM}} + E_{\text{vdW}}^{\text{QM-MM}} + E_{\text{bonded}}^{\text{QM-MM}} \quad (48)$$

where $E_{\text{ES}}^{\text{QM-MM}}$ is not calculated separately, but is included in the E_{QM} term in the electrostatic embedding (EE) approach. The Hamiltonian is

$$\hat{H} = \hat{H}_{\text{QM}} + \hat{H}_{\text{QM-MM}} \quad (49)$$

where \hat{H}_{QM} is the Hamiltonian for the QM region, and $\hat{H}_{\text{QM-MM}}$ is the Hamiltonian for the electrostatic interaction between QM system (electrons and nuclei) and MM system (modelled as point charges). $E_{\text{vdW}}^{\text{QM-MM}}$, which represents the van der Waals interaction (the dispersion interaction) and other short-range repulsive interactions, is normally described by a Lennard-Jones potential (Eq. 44). The last term is employed only when there are chemical bonds between the QM and MM regions, and it is calculated by the same force field as for E_{MM} . In this thesis, the hydrogen-link (HL) atom approach is used for the QM/MM boundary.

In our calculations, the total simulated system is divided into three subsystems: system 1 is the QM region that contains the active site of enzyme, and systems 2 and 3 are MM regions. System 2 typically consists of all residues within ~ 6 Å of system 1, and may be relaxed by MM during the geometry optimisation. System 3 is the remaining part and is kept fixed.

In our QM/MM approach, the energy is obtained from Eq. 50,

$$E_{\text{QM/MM}} = E_{\text{QM1+ptch23}}^{\text{HL}} + E_{\text{MM123,q}_1=0}^{\text{CL}} - E_{\text{MM1,q}_1=0}^{\text{HL}} \quad (50)$$

where $E_{\text{QM1+ptch23}}^{\text{HL}}$ is the QM energy of the QM region truncated by HL atoms and embedded in a set of point charges modelling systems 2 and 3. $E_{\text{MM1,q}_1=0}^{\text{HL}}$ is the MM energy of the QM system, still truncated by HL atoms, but without any electrostatic interactions. Finally, $E_{\text{MM123,q}_1=0}^{\text{CL}}$ is the classical energy of all atoms in the system with carbon link (CL) atoms and with the charges of the QM system set to zero (to avoid double counting of the electrostatic interactions).

2.4 Big-QM Approach

A of the problems with both the QM-cluster and QM/MM methods is that the energies vary strongly with the size of the QM region are selected. The big-QM approach was developed to obtain converged results.⁶⁸ In the big-QM calculations, all the important residues are included:⁶⁸ All chemical groups within 4.5–6 Å of the minimal QM system, all buried charged groups in the protein, and two capped amino acids round each residue in the minimal QM system are included in the QM calculations. This typically gives a QM system of 600–1200 atoms. All the big-QM calculations in the thesis were performed

on the coordinates from the QM/MM optimisation and with a point-charge model of surrounding because this gave the fastest calculations.⁶⁸

2.5 QTCP Method

A serious issue with QM/MM methods comes from local-minima problem, which is caused by the fact that a minimisation typically converges to the closest local minimum. There are no methods that always find the global minimum. In practice, it is not necessary that all groups are in their global minimum, but it is essential that they remain in the same local minimum throughout a reaction mechanism. A proper way to solve the problem is to calculate free energies, which involves sampling and averaging over all relevant thermally accessible structures.

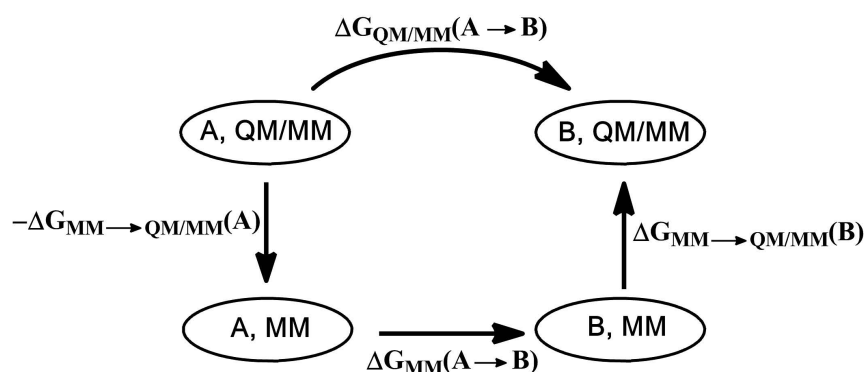


Figure 6. The thermodynamic cycle used in QTCP

In this thesis, the QM/MM thermodynamic cycle perturbation (QTCP) approach was used to calculate the free energy difference between two states at the QM/MM level of theory, using sampling only at the MM level.⁶⁹⁻⁷¹ The QTCP method employs the thermodynamic cycle in Figure 6. The top arrow is the QM/MM free energy from state A to state B, but it is computationally too expensive to calculate. An alternative way to obtain the QM/MM free energy is based on Eq. 51.

$$\Delta G_{\text{QTCP}}(A \rightarrow B) = \Delta G_{\text{MM}}(A \rightarrow B) + \Delta G_{\text{MM} \rightarrow \text{QM/MM}}(B) - \Delta G_{\text{MM} \rightarrow \text{QM/MM}}(A) \quad (51)$$

The first term, corresponding to the lower horizontal line in Figure 1, is the perturbation from state A to state B at the MM level. This energy can be calculated by free perturbation (FEP) according to

$$\Delta G_{MM}(A \rightarrow B) = -k_B T \ln \left\langle \exp \left(-\frac{E_A - E_B}{k_B T} \right) \right\rangle_A \quad (52)$$

where k_B is the Boltzmann constant, T is the temperature, and the angular brackets indicate an average over a MD ensemble, sampled for state A. One should test the precision of the result by calculating the same quantity while switching the A and B states.

The second and third terms are the free energy perturbation from MM to QM/MM description. The energy can be expressed as

$$\Delta G_{MM \rightarrow QM/MM}(X) = -k_B T \ln \left\langle \exp \left(-\frac{E_{QM1+ptch23}(X)}{k_B T} \right) \right\rangle_{MM} \quad (53)$$

where the X is either A or B. $E_{QM1+ptch23}$ was defined in Eq. 50.

3 Summary of the Articles

In this thesis, the six papers are grouped into two themes: Theoretical studies of [NiFe] hydrogenase (Papers I–IV) and of Mo-containing enzymes (DMSO reductase and formate dehydrogenase, Papers V and VI).

3.1 Paper I

The [NiFe] hydrogenases catalyse the reversible conversion of protons and electrons to H₂. However, the detailed reaction mechanism is unclear. The aim of paper I was to decide which of the Cys residues in the active site is most favourable to protonate. Therefore, the protonation states of the four Cys residues in four putative states in reaction mechanism, viz. the Ni-SI_a, Ni-R, Ni-C and Ni-L states, were studied in paper I.

In order to study the protonation states of the four Cys residues, a set of advanced methods were used: Geometries were optimised by the standard QM/MM approach with a small QM system; Accurate energies were calculated using the big-QM method (with 817 atoms in QM region), including all chemical groups within 4.5 Å of a minimal model of active site, all buried charged group, and moving the junctions at least two residues away from the active site. These calculations were performed at the TPSS/def2-TZVP level and energies were extrapolated to the B3LYP/def2-QZVPD level. To avoid the local-minima problem, QTCP calculations were also performed.

With these methods, we compared the energies of the species with different protonation states of the four Cys residues in the four putative intermediates in the reaction mechanism (Ni-SI_a, Ni-R, Ni-C and Ni-L) in reaction mechanism. Our results show that protonation of Cys-546 is most favourable for all four states, by 14–51 kJ/mol. For the Ni-R state, our results are consistent with a recent atomic-resolution crystal structure.⁴

3.2 Paper II

In paper II, we studied H₂ binding to the active site of [NiFe] hydrogenase. According to previous studies, both experimental and theoretical, two different binding modes have been suggested, viz. binding to Ni or binding to Fe. Moreover, the ground state of the Ni ion is not clear when H₂ binds to the active site. From previous theoretical studies, we know that the energies of this specific system are sensitive to the DFT methods and the size of the QM region. Thus, the results were different in different DFT calculations.

In order to solve the problems with different DFT methods and QM sizes, we employed CCSD(T), DMRG-CASPT2, as well as big-QM methods. The former two methods were used to calibrate the DFT methods, and the big-QM was used to avoid the QM-sizes problem. In the CCSD(T) calculations, a minimal model (18 atoms) was used because this method is computationally expensive. In the DMRG calculations, three models were used, with the active spaces of 22 electrons in 22 orbitals for the singlet state and 24 electrons in 24 orbitals for the triplet state (CAS(22,22) and CAS(24,24)). For the big-QM calculations, 819 atoms were included, and this size of the QM system is much larger than in any of the previous theoretical studies. All the geometry optimisations were performed by QM/MM method with the TPSS and B3LYP functionals, which allows us to evaluate which functional gives consistent results with CCSD(T) and DMRG.

Our results show that H₂ prefers to bind to the Ni ion in the singlet state, rather than to Fe, by at least 47 kJ/mol. For the triplet state, only H₂ binding to Fe species was found. In addition, we found that for this case, the TPSS functional gave better energies than B3LYP.

3.3 Paper III

In this paper, we have employed a new combined multiconfigurational and DFT method, CAS-srDFT, to explore the H₂ binding to the active site of the [NiFe] hydrogenase. The advantage of this method is that it captures dynamic correlation from DFT calculations, and the static correlation from the CASSCF wave functions. We know from paper II that H₂ prefers to bind to the Ni ion, and we therefore examined whether the CAS-srDFT calculations give the same conclusion or not.

Since the dynamic correlation is calculated by DFT methods, the better exchange–correlation functional is used, the more accurate the results will be. In this paper, we used the short-range PBE-based srPBE function by Goll et al. On the other hand, the range-separated framework is varied by the parameter μ . First, we studied how the energies changed when using different μ values. $\mu=0.4$ gave the best results, which is in agreement with the previous studies.⁶¹ The CASSCF method, used for the long-range part, is very sensitive to the orbitals in active space. Therefore, one must ensure that the orbitals are comparable in different calculations. We employed three different sizes of active space, viz. CAS(12,12), CAS(14,14), and CAS(16,16). In addition, three models were used to compare with paper II.

The CAS-srPBE results show that H₂ binding to Ni is more favourable than binding to Fe, which is consistent with the conclusion in paper II. For all three models, the effect of extending the active space from CAS(10,10) to CAS(14,14) was found to be small, ~ 2 kJ/mol. For model 1, we further employed CAS(16,16), which gave rise to a change of only 0.2 kJ/mol. Thus, the energies seem to converged with CAS(14,14). This is much smaller than in the previous DMRG calculations, which employed an active space of CAS(22,22), showing that the computational cost of CAS-srDFT is much lower, because a smaller active space can be used and the DFT calculation is much cheaper than the CASPT2 calculations.

3.4 Paper IV

Based on our results from paper I and II, we know that the Cys-546 is most easily protonated and that H₂ prefers to bind to Ni ion. However, the details of the reaction mechanism are still an open question, in particular whether the Ni-L state is involved in the reaction mechanism. In this paper, we have studied the full reaction mechanism of [NiFe] hydrogenase.

In this study, the QM/MM optimisations were carried out to obtain the geometries. More accurate energies were obtained by the big-QM calculations with 819 atoms. Moreover, DMRG-CASSCF calculations were carried out to study the electronic structures of the various states in the reaction mechanism.

Our calculations show that the Ni-L state is not involved in the reaction mechanism. Instead the Ni-C state is reduced by one electron and then the bridging hydride ion is transferred to the Cys-546 as a proton and the two electrons transfer to Ni ion. The cleavage of H–H bond is facile with an energy barrier of 23 kJ/mol based on our calculations. We also find that the reaction energies are sensitive to the size of QM system and the basis set, in agreement with our previous studies.

3.5 Paper V

In this paper, we have studied the effect of variations of the protein ligand in DMSO reductase. The DMSO reductase family is the largest and most diverse family of the mononuclear molybdenum oxygen-atom-transfer proteins. In the reaction cycle, the oxidation state of the Mo ion cycles between +IV and +VI. Remarkably, various members of the DMSO reductase family may employ three different protein-derived ligands (serine, cysteine, or selenocysteine). We have studied how the DMSO reductase reaction mechanism changes with alternative models of active site, varying the protein-derived ligand.

According to previous theoretical studies, the calculated barrier depends on the theoretical method and a proper account of dispersion and solvation effects is needed, together with large basis sets and accurate density functional theory (DFT) methods. In this paper, geometries were optimised in gas phase at the TPSS/def2-SV(P) level without any symmetry constraints. The energies were improved by single-point calculations using the B3LYP functional combined with the def2-TZVPD basis set. DFT-D3 dispersion corrections were applied to all single-point calculations and solvent effects were considered by the COSMO continuum solvent model with a dielectric constant of 4 to mimic the protein surrounding.

Our results show that the same mechanism was obtained with Ser, Cys, SeCys, OH⁻ and SH⁻ models for protein-derived ligands: The DMSO substrate first binds to Mo(IV) ion and then the S–O bond of DMSO is cleaved to generate product DMS. All five models gave similar activation barriers of 69–85 kJ/mol. However, with the doubly charged O²⁻ and S²⁻ models, the activation barriers were much higher, 212 and 168 kJ/mol, indicating that it is likely that the oxo and sulfido ligand are protonated to OH⁻ and SH⁻ during the reaction of enzymes employing these ligands.

3.6 Paper VI

Formate dehydrogenases (FDHs) catalyse the reversible conversion of formate to carbon dioxide. In paper VI, we have studied the reaction mechanism of Mo-containing formate dehydrogenase. In the previous experimental and theoretical studies, five putative mechanisms have been suggested. For these mechanisms, there are two important controversial questions: a) Does the cysteine in the active site dissociate from Mo during the reaction? and b) Does the substrate formate bind directly to Mo ion or not?

In the previous theoretical studies, small models were used and the protein surroundings were ignored. In this paper, the geometries were optimised with the QM/MM methods and the protein environment was considered by the MM method. Based on the QM/MM-optimised structures, we run big-QM calculations with 1121 atoms to obtain reliable energies. Moreover, thermal corrections from vibrational frequency calculations were added to the final energies.

Our results indicate that formate substrate does not bind directly to the Mo ion, but instead resides in the second coordination sphere. There, the sulfido group abstracts the hydride of formate, resulting in a Mo(IV)–SH state. Initially, the CO₂ product forms a thiocarbonate group with the Cys ligand. This step is quite favourable with an activation energy of 28 kJ/mol and a reaction energy of –39 kJ/mol. However, the CO₂ product is not released until the active site is oxidised by two electrons.

4 Conclusions and Outlook

In this thesis, we have studied three metalloenzymes, viz. [NiFe] hydrogenase, dimethyl sulfoxide reductase, and formate dehydrogenase. For [NiFe] hydrogenase, we have studied the protonation states of the four cysteine residues in active site, the H₂ binding site and the full reaction mechanism. This has allowed us to draw a clear picture of reaction mechanism. However, there still exist some unclear points. First, in the crystal structure, the glutamate acid, which forms a hydrogen bond to the terminal protonated cysteine in active site, is deprotonated, whereas in our calculations, the proton always moves to the glutamate acid from the cysteine. It would be interesting to study the crystal structure with theoretical methods, in order to calibrate our methods. Second, in our QM/MM studies, the big-QM approach was used to obtain stable QM energies. However, this approach is expensive and it introduces the problem to decide which residues have large effects on energy. Finally, why is the crystal structure of hydrogenase more similar to that of the triplet state, although our calculations indicate that this state is strongly unfavourable? We currently investigate these questions in our group.

For the Mo-containing enzymes, dimethyl sulfoxide reductase and formate dehydrogenase, we have studied the effect of variations in the protein-derived ligand for the former enzyme and the full reaction mechanism for the latter. We will continue to investigate enzymes that involve a Mo ion, e.g. nitrate reductases. The latter enzyme has an active site with the same Mo coordination sphere as formate dehydrogenase. However, it is still an open question whether the two enzymes follow a similar reaction mechanism. We currently study this enzyme in our group.

In paper III, we have studied the H₂ binding in the active site of [NiFe] hydrogenase with CAS-srDFT method. But, we did not consider the triplet state because open-shell calculations were not available at that time. However, now such code has been implemented in the DALTON package, which allows us to evaluate whether the CAS-srDFT method can handle the energy difference between the singlet and triplet states for this enzyme.

As DMRG can handle a large active space (~50 electrons in 50 orbitals), it can be used to calibrate DFT methods in some specific cases, e.g. when two or more metals are involved systems. It is also possible to study chemical reactions, e.g. bond cleavage, if the necessary orbitals are included in the active space. Currently, we investigate the lytic polysaccharide monooxygenases with CASSCF/CASPT2. However, an active space of CAS(16,16) or CAS(18,18) might be too small to produce comparable energies for reactant, transition and product states. Therefore, DMRG is a possible solution for this case.

Reference

- 1 A. Volbeda, M. H. Charon, C. Piras, E. C. Hatchikian, M. Frey and J. C. Fontecillacamps, *Nature*, 1995, **373**, 580.
- 2 Y. Higuchi, H. Ogata, K. Miki, N. Yasuoka and T. Yagi, *Structure*, 1999, **7**, 549.
- 3 W. Lubitz, H. Ogata, O. Ruediger and E. Reijerse, *Chem. Rev.*, 2014, **114**, 4081.
- 4 H. Ogata, K. Nishikawa and W. Lubitz, *Nature*, 2015, **520**, 571.
- 5 H. Ogata, T. Kramer, H. Wang, D. Schilter, V. Pelmeshnikov, M. van Gastel, F. Neese, T. B. Rauchfuss, L. B. Gee, A. D. Scott, Y. Yoda, Y. Tanaka, W. Lubitz and S. P. Cramer, *Nat. Commun.*, 2015, **6**, 7890.
- 6 R. Hidalgo, P. A. Ash, A. J. Healy and K. A. Vincent, *Angew. Chem. Int. Ed.*, 2015, **54**, 7110.
- 7 B. J. Murphy, R. Hidalgo, M. M. Roessler, R. M. Evans, P. A. Ash, W. K. Myers, K. A. Vincent and F. A. Armstrong, *J. Am. Chem. Soc.*, 2015, **137**, 8484.
- 8 S. Dementin, B. Burlat, A. L. De Lacey, A. Pardo, G. Adryanczyk-Perrier, B. Guigliarelli, V. M. Fernandez and M. Rousset, *J. Biol. Chem.*, 2004, **279**, 10508.
- 9 H. Ogata, Y. Mizoguchi, N. Mizuno, K. Miki, S. Adachi, N. Yasuoka, T. Yagi, O. Yamauchi, S. Hirota and Y. Higuchi, *J. Am. Chem. Soc.*, 2002, **124**, 11628.
- 10 A. Volbeda and J. C. Fontecilla-Camps, *Dalton Trans.*, 2003, 4030.
- 11 Y. Montet, P. Amara, A. Volbeda, X. Vernede, E. C. Hatchikian, M. J. Field, M. Frey and J. C. Fontecilla-Camps, *Nat. Struct. Biol.*, 1997, **4**, 523.
- 12 Metal Dihydrogen and σ -Bond Complexes: Structure, Theory and Reactivity; Kubas, G. J., Ed.; Kluwer Academic/Plenum Publishers: Dordrecht, Netherlands, 2001.
- 13 S. Qiu, L. M. Azofra, D. R. MacFarlane and C. Sun, *Phys. Chem. Chem. Phys.*, 2018.
- 14 M. Bruschi, M. Tiberti, A. Guerra and L. De Gioia, *J. Am. Chem. Soc.*, 2014, **136**, 1803.
- 15 R. Hille, J. Hall and P. Basu, *Chem. Rev.*, 2014, **114**, 3963.
- 16 R. Hille, *Chem. Rev.*, 1996, **96**, 2757.
- 17 J. L. Li, R. A. Mata and U. Ryde, *J. Chem. Theory Comput.*, 2013, **9**, 1799.
- 18 A. L. Tenderholt, J. J. Wang, R. K. Szilagy, R. H. Holm, K. O. Hodgson, B. Hedman and E. I. Solomon, *J. Am. Chem. Soc.*, 2010, **132**, 8359.
- 19 E. Hernandez-Marin and T. Ziegler, *Can. J. Chem.*, 2010, **88**, 683.
- 20 J. P. McNamara, I. H. Hillier, T. S. Bhachu and C. D. Garner, *Dalton Trans.*, 2005, 3572.

- 21 J. P. McNamara, J. A. Joule, I. H. Hillier and C. D. Garner, *Chem. Commun.*, 2005, **2**, 177.
- 22 A. Thapper, R. J. Deeth and E. Nordlander, *Inorg. Chem.*, 2002, **41**, 6695.
- 23 C. E. Webster and M. B. Hall, *J. Am. Chem. Soc.*, 2001, **123**, 5820.
- 24 A. Thapper, R. J. Deeth and E. Nordlander, *Inorg. Chem.*, 1999, **38**, 1015.
- 25 J. L. Li, M. Andrejic, R. A. Mata and U. Ryde, *Eur. J. Inorg. Chem.*, 2015, 3580.
- 26 J. L. Li and U. Ryde, *Inorg. Chem.*, 2014, **53**, 11913.
- 27 S. Metz and W. Thiel, *Coord. Chem. Rev.*, 2011, **255**, 1085.
- 28 M. Hofmann, *Inorg. Chem.*, 2008, **47**, 5546.
- 29 A. L. Tenderholt, K. O. Hodgson, B. Hedman, R. H. Holm and E. I. Solomon, *Inorg. Chem.*, 2012, **51**, 3436.
- 30 A. M. Appel, J. E. Bercaw, A. B. Bocarsly, H. Dobbek, D. L. DuBois, M. Dupuis, J. G. Ferry, E. Fujita, R. Hille, P. J. A. Kenis, C. A. Kerfeld, R. H. Morris, C. H. F. Peden, A. R. Portis, S. W. Ragsdale, T. B. Rauchfuss, J. N. H. Reek, L. C. Seefeldt, R. K. Thauer and G. L. Waldrop, *Chem. Rev.*, 2013, **113**, 6621.
- 31 L. B. Maia, J. J. G. Moura and I. Moura, *J. Biol. Inorg. Chem.*, 2015, **20**, 287.
- 32 T. Hartmann, N. Schwanhold and S. Leimkuhler, *Bba-Proteins Proteom*, 2015, **1854**, 1090.
- 33 M. Jormakka, B. Byrne and S. Iwata, *Curr. Opin. Struct. Biol.*, 2003, **13**, 418.
- 34 P. Schrapers, T. Hartmann, R. Kositzki, H. Dau, S. Reschke, C. Schulzke, S. Leimkuhler and M. Haumann, *Inorg. Chem.*, 2015, **54**, 3260.
- 35 H. C. A. Raaijmakers and M. J. Romao, *J. Biol. Inorg. Chem.*, 2006, **11**, 849.
- 36 H. Raaijmakers, S. Macieira, J. M. Dias, S. Teixeira, S. Bursakov, R. Huber, J. J. G. Moura, I. Moura and M. J. Romao, *Structure*, 2002, **10**, 1261.
- 37 J. C. Boyington, V. N. Gladyshev, S. V. Khangulov, T. C. Stadtman and P. D. Sun, *Science*, 1997, **275**, 1305.
- 38 M. Jormakka, S. Tornroth, B. Byrne and S. Iwata, *Science*, 2002, **295**, 1863.
- 39 L. B. Maia, I. Moura and J. J. G. Moura, *Inorg. Chim. Acta* 2017, **455**, 350.
- 40 M. Born and R. Oppenheimer, *Ann. Phys.*, 1927, **84**, 457.
- 41 D. R. Hartree, F. R. S. Hartree and W. Hartree, *Proc. R. Soc. Lond. A Math. Phys. Sci.*, 1935, **150**, 0009.
- 42 V. Fock, *Z. Angew. Phys.*, 1930, **61**, 126.
- 43 D. R. Hartree, *P. Camb. Philos. Soc.*, 1928, **24**, 89.
- 44 F. Jensen, *Introduction to Computational Chemistry*; Wiley: New York, 2013.
- 45 J. C. Slater, *Phys. Rev.*, 1930, **36**, 0057.

- 46 S. F. Boys, *Proc. R. Soc. Lon. Ser. A*, 1950, **200**, 542.
- 47 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297.
- 48 R. Ditchfield, W. J. Hehre and J. A. Pople, *J. Chem. Phys.*, 1971, **54**, 724.
- 49 P. O. Widmark, P.-Å. Malmqvist and B. O. Roos, *Theor. Chim. Acta*, 1990, **77**, 291.
- 50 C. Moller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 0618.
- 51 D. Maurice and M. Head-Gordon, *Mol. Phys.*, 1999, **96**, 1533.
- 52 M. Head-Gordon, R. J. Rico, M. Oumi and T. J. Lee, *Chem. Phys. Lett.*, 1994, **219**, 21.
- 53 J. Cizek, *J. Chem. Phys.*, 1966, **45**, 4256.
- 54 M. W. Schmidt and M. S. Gordon, *Annu. Rev. Phys. Chem.*, 1998, **49**, 233.
- 55 B. O. Roos, *Int. J. Quantum Chem.*, 1980, **17**, 175.
- 56 K. Andersson, P.-Å. Malmqvist and B. O. Roos, *J. Chem. Phys.*, 1992, **96**, 1218.
- 57 S. R. White, *Phys. Rev. Lett.*, 1992, **69**, 2863.
- 58 W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, 1133.
- 59 S. F. Sousa, P. A. Fernandes and M. J. Ramos, *J. Phys. Chem. A*, 2007, **111**, 10439.
- 60 A. Savin, *Recent Developments and Applications of Modern Density Functional Theoretical*. Elsevier: Amsterdam, 1996, pp 327.
- 61 E. Fromager, J. Toulouse and H. J. A. Jensen, *J. Chem. Phys.*, 2007, **126**.
- 62 T. Leininger, H. Stoll, H. J. Werner and A. Savin, *Chem. Phys. Lett.*, 1997, **275**, 151.
- 63 J. Heyd and G. E. Scuseria, *J. Chem. Phys.*, 2004, **120**, 7274.
- 64 J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**, 8207.
- 65 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
- 66 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227.
- 67 U. Ryde, *Method Enzymol*, 2016, **577**, 119.
- 68 L. Hu, P. Söderhjelm and U. Ryde, *J. Chem. Theory Comput.*, 2013, **9**, 640.
- 69 T. H. Rod and U. Ryde, *J. Chem. Theory Comput.*, 2005, **1**, 1240.
- 70 T. H. Rod and U. Ryde, *Phys. Rev. Lett.*, 2005, **94**, 138302.
- 71 V. Luzhkov and A. Warshel, *J. Comput. Chem.*, 1992, **13**, 199.

Acknowledgements

Time flies! I have spent almost four years at Lund University. I remember four years ago when I was in China and I was so excited that I can go to Sweden, but at the same time I was also afraid because I did not know how the future life would be. Now, I can give me the answer: I really enjoyed my time here and I have learnt a lot both in research and in life. Here, I am very grateful to you, my friends, and without you I could not have such a great experience.

First of all, I would like to thank my supervisor, Ulf Ryde. You are always patient and positive when we discuss our research, and even when I made some stupid mistakes. This means a lot to me. You help me much, not only in my research, but also in my life. I also would like to thank my co-supervisor, Ebbe Nordlander for your inspiring discussions. Although we did not meet very often, every time I talk with you, I get some new ideas for my research.

I am thankful to everyone in my group, Paulius, Francesco, Martin, Majda, Lili, Octav, Erik and all visitors. When I came to KC my first time, Paulius brought me to the student centre to help me to apply for the entrance card. Also, you taught me how to install some software on Linux system. Francesco, you were always nice when I asked you some questions and helped me a lot in my research; I miss you. Martin, thank you that you told me a lot about Sweden, and thank you for your help at the programing course. Majda, thank you in many aspects, courses, research, travel and so on. Lili, thank you for your many advices on research and life, e.g., how to treat people, how to improve my appearance, etc. Octav, you are so clever, you know a lot and learn fast, and you help me a lot in my research and courses. Erik, thank you for teaching me DMRG, CAS, and srDFT calculations. I am very happy that I can collaborate with you in my research. I thank all the visitors, Azar, Adrian, Christine, Nadja, Fatemeh, Casper, Adam, Alfonso, Melanie, Meiting, and Anna. I also thank Pär, Soumendranath and Esko. I really enjoy the discussions in our journal club.

I am thankful to Kristine, Quan and Simon. I had a nice experience because of you when I was in Leuven. I learnt a lot about DMRG and CCSD(T). Also, I miss that time when we played bowling together.

I would like to thank the DMRG group in our department, Valera, Per-Åke, Martin, and Erik. I learnt a lot from you.

I am thankful to Fredrik, my officemate during my first year in KC. You helped me a lot, in both research and daily life.

I am very grateful to my Chinese friends. Wei, you are my first Chinese friend I met in Sweden. I clearly remember it was at night around 11 o'clock on 15th September in 2014. I called you because I was so hungry, and you brought some bread and apples. In the next day, you brought me to KC. Around 11 in the morning, Fei came to me: "I heard that a Chinese student came here in Ulf's

group”, Fei said to me. Then, you introduced me to Ruiyu. You told me how to get an apartment, how to register a personal number and so on. Days after, I met Weimin at Annil’s party, and then I learnt YouTube from you. This was the first time I heard about YouTube. After that I knew what to do when I was at home. Months after, Hongduo came back from maternity leave, and we had a long talk about how to enjoy my stay here in Sweden. From then on, I met many Chinese friends, Xiaoting, Feifei, Junsheng, Wei Dang, Delin, Bin, Ke, Meina, Lingdong, Qianjin, Junhao, Fenying ... I have a lot friends now in Lund more than I can fit here. I appreciate you all very much and because of you, my life becomes more and more beautiful.

I would like to thank Ingrid, Helena, and Maria for taking care of us. I remember the first day in KC (the second day in Sweden), Ingrid, you helped me to register my PhD. You are so nice and patient. Helena and Maria, you help me a lot when I had some problems at KC.

Finally, I would like to thank my families. My parents, you always support me during the four years. You always said “take care of yourself, your father and me are fine, and do not worry about us”. I could not go further without your supporting. I also thank my elder sister. Because you take care of our parents, I can study in Sweden. My dear wife, Xiaoqiong, thank you for your company in my life, I love you forever.

Thank you very much!
谢谢!

Geng Dong
东庚