SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of Biotechnology and Biosciences

PhD program in Biology and Biotechnology Cycle XXX

Computational modelling of macromolecules: prediction of the 3-D structure, catalytic activity and dynamic features of proteins

Jacopo Vertemara

Registration number 717171

Tutor: Professor Luca De Gioia

Coordinator: Professor Marco Ercole Vanoni

ACADEMIC YEAR 2016/2017

Summary

Scope of Biotechnology is the use of biological constituents (systems or organisms) to make products useful to the mankind. A way to achieve this goal is to implement enzymes in the productive process. The use of enzymes allows a lot advantages: mild productive conditions in term of temperature and pressure with a great economic saving, a reduce use of toxic reagents and low production of polluting wastes, possibility to implement reactions which are difficult to realize without enzymes (i.e. stereospecific and stereoselective reactions). Unfortunately, the implementation of enzymes is not easy to accomplish because often proteins are unstable out from natural organism or under productive conditions, moreover the use of non-natural substrates can reduce the catalytic efficiency. A possible solution to overcome these problems is to use mutated enzymes. In order to design a specific mutation with a desired effect it is essential to know some key aspects of protein like 3D structure, dynamic features and the catalytic pathway of the reaction. In medicine a relevant goal of Biotechnology is to characterize proteins in diseases-related condition in order to point out the molecular determinants of specific pathology and to develop new and more efficient therapeutic strategies. All these aspects can be investigated with computational approaches which often represent the only way to have molecular insights of such biological systems. In this work I used different computational techniques, in particular QM and MM methods, to study the relation between structure and activity.

The first project was the characterization under structural point of view of the FeMo-cofactor presents in the active site of nitrogenase. My work was focused on the structural characterization of the key cofactor state E_4, that is now accepted as the intermediate that bind N2 in the active site. For this purpose, DFT simulations have been performed on the FeMo-co active site considering all the iron oxidation state assignments proposed in literature ($[5Fe^{III} : 2Fe^{II}]$; $[3Fe^{III} : 4Fe^{II}]$; $[1Fe^{III} : 6Fe^{II}]$) with protonated and not protonated forms of homocitrate. Only in cases of $[3Fe^{III} : 4Fe^{II}]$ with protonated form of homocitrate the optimized E_4 structure shows two hydrides bridged to the same iron atoms (Fe_6 and Fe_7) in an orthogonal fashion, consistently with experimental data. I have also studied the electronic properties of E_4 intermediate in order to evaluate the spin alignment of ferrous and ferric sites, with peculiar focus on the two Fe ions sharing hydride ligands. DFT broken symmetry calculations showed a parallel spin-coupling pattern for the two iron atoms ($\downarrow Fe_6$: $\downarrow Fe_7$).

Another issue investigated is the effect of the substitution of $Arg_{96} \rightarrow Gln$ on the resting state of the MoFe protein (E_0). Experimental evidences show that in presence of this mutation the active site of nitrogenase is able to (non-covalently) bind acetylene even in the resting state (whereas WT protein cannot). In order to rationalize the observed behavior, DFT approach has been used to evaluate the binding energy between acetylene and FeMo-co active site in the cases of mutant and wild type systems. In line with experiments, the binding energy results more favourable for the mutant than for the wild type enzyme. An explanation can be found in the different chemico-physical properties of the region available for acetylene binding in cofactor proximity: in presence of the substitution $Arg_{96} \rightarrow Gln$ the pocket environment becomes more hydrophobic than in the wild type case. Considering the nonpolar nature of acetylene, the mutated MoFe protein has therefore active site features that make it more suitable for substrate binding.

The second project was the characterization of the hydrolytic mechanism of a *de novo* design peptide TRIL9CL23H (an analogue of carbonic anhydrase) trough DFT calculations and identification of a possible binding site for substrates using flexible docking. The mechanism proposed for TRIL9CL23H is a merge of α-anhydrase and β-anhydrase ones: as in α-anhydrase there is a zinc coordinated by three histidine residues and a water molecule, as in β-anhydrase a glutamate residue accepts the water's proton from active site. Contrary to carbonic anhydrase, results show that the rate determining step is the release of the products from active site and not the activation of water. According to docking results the binding site is located on the side of proteins close to the active site.

The last project has investigated the effect of the R10T amino acid change in the Mre11 protein (a subunit of MRX complex), which causes a rapid DSB resection. The evolutionarily conserved MRX complex plays a central role in the cellular response to DSBs, as it is implicated in controlling end resection and in maintaining the DSB ends tethered to each other. Given the importance of end-resection in DSB repair and in preventing oncogenic rearrangements, proteins involved in this process might be suitable targets to selectively increase the chemo-/radio-sensitization of tumor cells. This approach needs mechanistic insights into the MRX complex and the understanding of how its subunits work at DSBs. For this purpose, the wild type and the mutated dimer of Mre11 were studied trough molecular dynamics simulations. Results pointed out that the R10T substitution alters the mobility of the capping domain of Mre11, movement that is implicated in DNA unwinding. This mutation augments the rotation of the capping domain that can lead to a better DNA unwind activity. In this way nucleases that are involved in the DSB resection can have a better access to the DNA.

Riassunto

Uno degli obbiettivi delle Biotecnologie è quello di sfruttare sistemi biologici (cellule intere o enzimi) per la produzione di prodotti utili alla collettività. In questo senso sta prendendo sempre più piede l'utilizzo di enzimi all'interno dei processi produttivi industriali. L'utilizzo di enzimi comporta molti vantaggi tra cui: l'impiego di condizioni di produzione più blande in termini di temperatura e pressione con un conseguente risparmio economico e di energia; diminuzione nell'utilizzo di composti tossici e nella produzione di scarti inquinanti; possibilità di effettuare reazioni stereospecifiche e stereoselettive. Sfortunatamente l'impiego di enzimi nei processi produttivi non è sempre facile poichè tali enzimi potrebbero risultare instabili e/o incapaci di processare substrati non naturali. Un possibile modo per ovviare a questi problemi è quello di utilizzare enzimi modificati. Per poter progettare delle mutazioni che apportino specifiche proprietà ad un enzima è fondamentale conoscere la struttura 3-D, le proprietà dinamiche e catalitiche dell'enzima stesso. Un altro ambito dove le Biotecnologie ricoprono un ruolo importante è in campo medico per studio di malattie complesse come il cancro. L'obbiettivo è quello di caratterizzare a livello molecolare il ruolo che proteine mutate hanno nel progredire di una data malattia in modo tale da poter sviluppare cure sempre più specifiche ed efficaci. Tutto ciò può essere studiato tramite approcci computazionali (in alcuni casi tali metodi rappresentano la sola via percorribile). Durante il mio PhD ho utilizzato diverse tecniche computazionali, in particolare metodi QM e MM, per studiare la relazione struttura/attività delle proteine. Il primo progetto consiste nella caratterizzazione strutturale del cofattore metallico FeMo-co presente nel sito attivo delle Mo-nitrogenasi. L'obbiettivo è studiare da un punto di vista strutturale l'intermedio E_4 che è attualmente ritenuto essere lo step nel quale avviene il legame con N_2. A tale scopo sono state svolte simulazioni DFT sul FeMo-co considerando ogni possibile stato redox dei ferri proposto in letteratura ($[5Fe^{III} : 2Fe^{II}]$;

$[3Fe^{III} : 4Fe^{II}]$; $[1Fe^{III} : 6Fe^{II}]$) e la forma protonata e non protonata dell'omocitrato. Solo con la configurazione $[3Fe^{III} : 4Fe^{II}]$ e con l'omocitrato protonato la struttura ottimizzata di E_4 mostra la presenza di due idruri a ponte disposti in modo ortogonale tra loro e legati entrambi ai medesimi atomi di ferro (Fe_6 e Fe_7) in accordo con i dati EPR. Tramite Broken Symmetry si è studiato anche il corretto allineamento di spin dei vari atomi di ferro con particolare riguardo per i due ferri (Fe_6 e Fe_7) che legano gli idruri. Un altro aspetto studiato riguardante le Mo-nistrogenasi è stata la caratterizzazione della mutazione $Arg_{96} \rightarrow Gln$. Tale sostituzione rende possibile il binding dell'acetilene al sito attivo delle nitrogenasi anche durante il resting state E_0. Per tentre di dare una spiegazione a tale fatto, si è calcolata l'energia di binding dell'acetilene sia per sito attivo wild type che per quello mutato tramite simulazioni DFT. L'energia di binding dell'acetilene risulta essere più favorevole per il mutante che per il wild type in linea con i dati sperimentali. Dalle simulazioni effettuate risulta che la sostituzione $Arg_{96} \rightarrow Gln$ altera le proprietà chimico-fisiche della tasca di legame, in particolare tale mutazione ha l'effetto di aumentare le caratteristiche idrofobiche del sito di binding. Considerando la natura apolare dell'acetilene, l'aumento del carattere idrofobico dell'intorno aminoacidico del FeMo-co spiega bene il motivo del binding anche durante il resting state E_0.

Il secondo progetto è la caratterizzazione del meccanismo catalitico di un enzima di *de novo* design tramite simulazioni DFT e di docking molecolare. La proteina presa in considerazione è TRIL9CL23H, un analogo dell'anidrasi carbonica. Il meccanismo catalitico proposto per TRIL9CL23H risulta essere una combinazione dei meccanismi catalitici della α-anidrasi e della β-anidrasi: come accade nel α-anidrasi il sito attivo è costituito da una molecola di acqua coordinata ad un atomo di zinco, a sua volta coordinato con geometria tetraedrica a tre istidine; come accade invece nella β-anidrasi l'accettore del protone perso dalla molecola d'acqua coordinata allo zinco risulta essere un glutammato. I risultati mostrano che a differena dell'anidrasi carbonica, dove il passaggio limitante è l'attivazione della molecola d'acqua, nel caso di TRIL9CL23H il collo di bottiglia di tutto il processo catalitico risulta essere l'uscita dei prodotti dal sito attivo dell'enzima.

Nell'ultimo progetto si è studiato l'effetto della sostituzione R10T della subuità Mre11 del complesso MRX (Mre11-Rad50-Xrs2). Il complesso MRX svolge un ruolo importante nella regolazione della risposta da rottura della doppia elica di DNA (Double Strand Break) poichè è responsabile della scelta tra Non Homologous End Joining (NHEJ) e

ricombinazione omologa (HR) per la sua riparazione. Dai dati sperimentali è emerso che la variante Mre11-R10T è capace di bypassare sia la mancanza della proteina Sae2 e sia di indurre una più alta attivita di resection. L'obbiettivo di questo studio è quello di caratterizzare l'effetto di questa sostituzione utilizzando tecniche di simulazione di dinamica molecolare (MD). L'analisi delle dinamiche ha mostrato che la sostituzione R10T provoca una diversa disposizione spaziale del capping domain di Mre11 che ha come effetto quello di aumentare l'attività di svolgimento del DNA. Questa aumentata predisposizione di Mre11-R10T ad aprire la doppia elica di DNA garantendo così alle nucleasi un più facile accesso, spiega l'aumentata attività di resection osservata sperimentalmente.

Contents

Chapter 1

Introduction to computational approaches

The simplest definition of biotechnology is a technology that uses living organisms or their components (e.g. enzymes) to develop or make products in order to improve the quality of human life. In medicine, for instance, a relevant goal of Biotechnology is to characterize proteins in diseases-related condition in order to point out the molecular determinants of specific pathology and to develop new and more efficient therapeutic strategies. Nowadays one of the most studied diseases is cancer. Every year more than 3 million people in Europe are diagnosed with cancer, a disease that is caused by alterations of genome as well as the proteome. The major goal of the biotechnological research is to identify the mechanisms that allow cancer to evade normal cellular control checkpoints in order to develop effective diagnostic, treatment and prevention regimes. This goal can be achieved by studying at molecular level the interactions developed among proteins inside a cell. Other great effort of biotechnologies is to implement enzymes in productive processes. The usage of enzymes provides a lot of advantages in economic and production terms for industry. In fact, it is possible to develop processes that involve mild temperatures and mild pressures with a great energy and economic saving, to reduce the use of toxic reagents and so to reduce the polluting wastes, and finally to implement easily stereospecific and stereoselective reactions. Unfortunately, there are also disadvantages in usage of enzymes because they might be unstable out from natural host or under productive conditions, moreover it is not guaranteed that standard enzymes were able to work with non-natural substrates. Bioengineering tries to overcome such problems by producing a "new version" of an enzyme giving it new properties. To design specific mutations with the aim to introduce a novel property into a protein, the structure and the catalytic features of that protein must be note. Furthermore, it is also necessary to test the effect of such mutations and this step is often expensive and time consuming. Computational approaches can be helpful in this sense, because they provide a faster and cheaper way to investigate structure and chemical properties of a molecule than experimental process. Computer based techniques can be sorted into two main classes: the quantum mechanical (QM) methods and molecular mechanics (MM) methods. QM and MM methods differ by the resolution scale that use to address issues concerning a biological system. The first one uses as "building blocks" atomic nuclei and electrons and it stems from quantum mechanics, the second one describes each atom as a whole and it is based on Newton (classic) physics. It follows that the QM methods are useful to study chemical and physical properties of a chemical system whereas the MM methods are able to describe the dynamic features of macromolecules. The present dissertation reports four case studies in which it will be disclosed both usefulness and limits of either approaches with the aim to provide a guideline to investigate biotechnologically relevant systems.

1.1 Quantum mechanical approach

Computational quantum chemistry is a technique that provides a detailed description of a chemical reaction including elusive transition states and short-lived intermediates that are impossible to investigate with experimental techniques. Quantum chemistry applies the fundamental laws and equations of quantum mechanics to study chemical systems. More in detail, quantum chemical methods stem from the Schrödinger equation that treats molecules as collections of nuclei and electrons, without any reference to their chemical bonds. By solving the non-relativistic time-independent Schrödinger equation

$$E(\psi) = \hat{H}(\psi)$$

where $\hat{H}$ is the electronic Hamiltonian operator, E the energy of the molecule and ψ the wavefunction, which describes the position of electrons and nuclei within the molecule, it is possible to obtain the energy of the system represented by the wavefunction itself. The wavefunction ψ represents the configurations of the electrons in the molecular orbitals of a certain molecule. As aforementioned, by applying the Hamiltonian operator to the wavefunction it is possible to calculate the energy of the system. The Hamiltonian operator is evaluated as follows:

$$\hat{H} = -\sum \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum \frac{\hbar^2}{2m_k}\nabla_k^2 - \sum_i\sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{i<j} \frac{e^2}{r_{ij}} + \sum_{k<l} \frac{e^2 Z_k Z_l}{r_{kl}}$$

where the five terms are the kinetic energy of the electrons, the kinetic energy of the nuclei, the coulomb attraction between electrons and nuclei, the repulsion between electrons and the repulsion between nuclei, respectively. Unfortunately, the Schrödinger equation can be solved exactly only for the simplest mono-electronic systems such as the hydrogen atom. In order to resolve such equation also in case of more complex multi-electronic systems is necessary to introduce some approximations.

The first one is the Born-Oppenheimer approximation in which the coupling between nuclei and electrons motions are neglected. Due to the large difference in masses between electrons and nuclei (recall that protons and neutrons are about 2000 times heavier than electrons), the electronic relaxation is instantaneous with respect to nuclear motion. Thanks to this expedient, the Hamiltonian lacks the nuclear kinetic energy and the nuclear-nuclear potential energy term is considered constant for a given geometry.

The second approximation is the Hartree-Fock (HF) approximation where each electron is described by an orbital and the total wave functions is given as a product of orbitals. This simplification is obtained by using the independent-particle model that considers the motion of one electron to be independent of the dynamics of all other electrons. Thus, molecular orbitals (MOs) and molecular orbital energies are derived using a one-electron formalism and then the energy of multi-electron system can be determined as the sum of energies of the occupied one-electron orbitals. Since electrons are indistinguishable one to another (electrons are fermions with spin $= 1/2$) the total wave function must be anti-symmetric (the permutation of two electrons causes the sign of the wave function to change) in agreement with the Pauli principle. The anti-symmetry of the wave function is achieved by adapting the orbitals in a Slater determinant.

$$\Psi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \cdots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \cdots & \chi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(N) & \chi_2(N) & \cdots & \chi_N(N) \end{vmatrix}$$

HF wave-function is a single-determinant one which means the electron-electron interactions are taken into account in average fashion and consequently the instantaneous correlation between electrons are neglected. Despite this approximation, the HF method is able to well describe the molecular systems proximal to equilibrium geometry. The Hartree-Fock methodologies are classified into two classes depending on how α spin-orbitals and β spin-orbitals are treated. The Restricted Hartree-Fock (RHF) uses a single function twice with opposite sign to describe α and β electrons, on the contrary, Unrestricted Hartree-Fock (UHF) uses different functions for α and β orbitals. Usually RHF is implemented to study closed-shell systems with all electrons being paired, whereas the UHF is applied for open-shell systems where the number of α and β electrons are not equal.

A third approximation is the introduction of a basis set. As aforementioned, the total HF wave function is expressed as a Slater determinant formed from the single occupied molecular orbitals (MOs). MOs are constructed as a linear combination of a finite set of functions known as basis set. Different types of basis set are available and the choice of appropriate one is a crucial step in quantum chemical calculations. The main point is to choose a basis set that reproduces the physical and chemical properties of the system and that ensures fast convergence. The process that leads to the convergence of the MOs energy is the iterative self-consistent field (SCF) method. This process involves the following iterative steps:evaluation of wave functions for all the occupied MOs to construct the one-electron Hamiltonians, usage of such Hamiltonians to provide a new set of wave functions different from the previous ones. The process is repeated until the difference between newly determined wave functions and preceding wave functions falls below an arbitrarily set threshold.

Density-functional theory (DFT) is an independent-particle model like HF and, even if it is comparable to HF method under a computational point of view, it provides significantly better results. DFT is based on Hohenberg-Kohn theorems that demonstrate how ground state associated electron density allows the calculation of a certain number of properties of the ground state itself. The ground state energy of the system E_0 is a functional (a prescription for producing a number from a function, which in turn depends on variables) of the electron density

$$E_0 = E_0(\rho)$$

The Kohn-Sham theory states that the ground-state electronic energy of a molecular system can be calculated summing electrons kinetic energy, nuclear-electron potential energy, electron-electron coulomb energy and exchange-correlation energy. The exchange-correlation energy is a functional of electron density that involves both exchange effects, the probability of finding two electrons with the same spin in the same point, and correlation effects, the probability of finding two electrons with opposite spin.

1.2 Force field methods

As told at the onset of the previous paragraph, solving the Schrödinger equation provides a complete description of the energy of a chemical system. In principle this equation can be applied to calculate the energy of all molecules, but actually its implementation for huge systems (i.e. proteins) is prohibitive due to the high complexity of *ab initio* techniques. *Ab initio* techniques are computationally very expensive and the more atoms, the longer time is needed for the calculation. Larger chemical systems are studied using the force field method also referred as molecular mechanics (MM) method. MM methods use as "building block" the whole atom (electrons are not considered) and demand to provide explicitly the bonding information contrary to QM methods. In particular, atoms are described just like sort of "balls" having different size and softness whereas bonds are "springs-like" having different lengths and stiffness. MM theory is based on the application of Newton's equation:

$$F_i(t) = m_i a_i(t)$$

where F_i is the force acting on i^{-th} atom at t, m_i is the mass and a_i is the acceleration. For time-independent phenomena, the problem is limited to calculating the energy at a given geometry. Potential energy is expressed as a sum of terms each describing the energy required for distorting a molecule in a specific fashion:

$$U = E_{str} + E_{bend} + E_{tors} + E_{wdw} + E_{el}$$

where E_{str} represents the energy function for stretching a bond, E_{bend} represents the energy required for bending an angle, E_{tors} is the torsional energy for rotation around a bond, E_{vdw} represents the non-polar interaction energy and E_{el} represents the electrostatic energy. Both E_{str} and E_{bend} are described by Hook's law that explains the behavior of elastic bodies in the presence of a given external force. The harmonic approximation of the E_{str} and E_{bend} works only if the system is close to equilibrium. In other words, E_{str} takes into account the energy associated with vibrations around bond length at equilibrium, whereas E_{bend} accounts for the energy associated with little variations around the valence angle at equilibrium. E_{tors} describes the energy associated with a torsion, that corresponds to the rotation around B-C bond considering four atoms A-B-C-D. Due to the fact that after a rotation of 360° the energy returns to the same value, E_{tors} is described by a periodic function. The last two terms E_{vdw} and E_{el} describe the non-bonding interactions between atoms: these interactions represent the repulsion or attraction between all couples of atoms that are not directly bonded. E_{vdw} is represented by a Morse potential, a function that well describes the energy variation of two interacting atoms. The interaction energy is zero when the two atoms are at infinite distance, whereas increases when the atoms come closer to a minimum energy point. E_{vdw} takes into account the effects of dispersion forces, (London interactions) and induced dipole-induced dipole interactions. E_{el} is the Coulomb potential that describes the energy interaction between two charged particles.

By solving the classical equation of motion, the second Newton's law aforementioned in the section, MM methods can give information on the system dynamics in the atomic scale by a technique referred to as MM-MD (molecular mechanical molecular dynamics). From the knowledge of the forces acting on each atom using the potential energy function, it is possible to determine the acceleration of each atom of the system. By integrating the equation of motion, it is possible to obtain a trajectory that describes the evolution of the system over time. Here is reported the mathematical formalism

$$F_i = m_i \frac{d^2\vec{r}_i(t)}{dt^2}$$

where F_i is the forces acting upon i^{-th} particle at time t, mi is the mass of particle i^{-th} and $\vec{r}_i(t)$ is the position vector of particle i^{-th}. The integration of this second-order differential equation is a hard task due to the many-body nature of the problem, thus it is discretized and solved numerically. Aim of the numerically integration is to find an

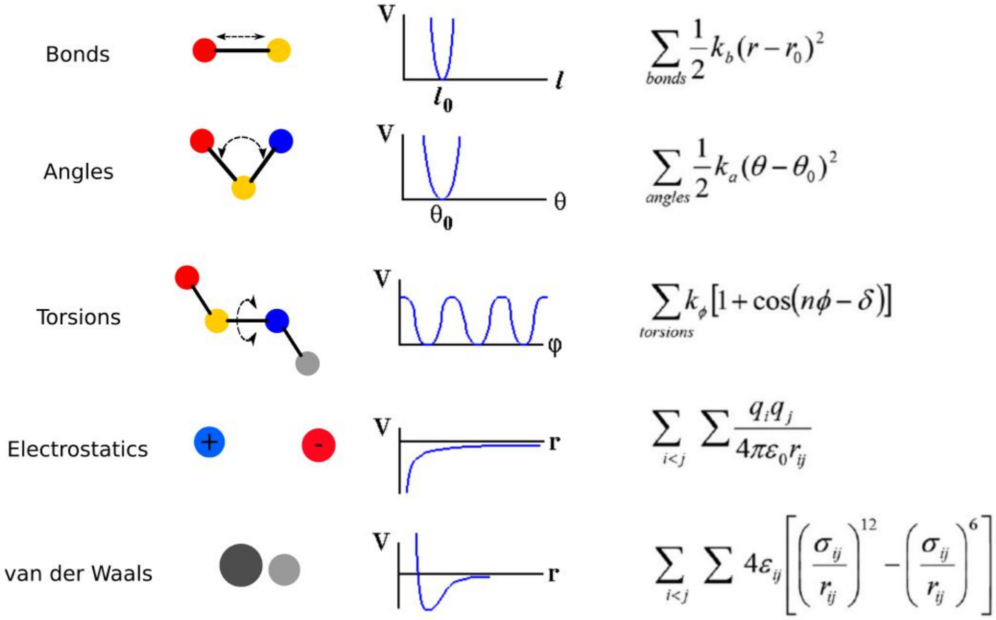


Figure 1.1: Illustration of the fundamental force field energy terms.

expression that defines positions $\vec{\boldsymbol{r}}_i$(t+Δt) at time t+Δt in terms of the already known positions at time t. This task is accomplished by the Verlet algorithm

$$\vec{\boldsymbol{r}}_i(t + \Delta t) \approx 2\vec{\boldsymbol{r}}_i(t) + \vec{\boldsymbol{r}}_i(t - \Delta t) + \frac{F_i(t)}{m_i}\Delta t^2$$

It is crucial that Δt has to be smaller than the time of the fast motions in the system. In the case of biological macromolecules the fast motions are the vibrations of light atoms (i.e. O-H bond) that occur in several femtosecond implying that Δt should be on a sub-femtosecond scale to ensure stability of the integration (a classical time step is 2fs). The initial velocities of particles are usually randomly chosen from a Maxwell-Boltzmann at a given temperature

$$f(\upsilon) = N\sqrt{\left(\frac{m}{2\pi k_B T}\right)^3} e^{-\frac{m\upsilon^2}{2k_B T}}$$

where k_B in the Boltzmann's constant, T is the thermodynamic temperature. m is the particle mass, N the number of the particles and υ is the velocity of the particle.

1.3 Potential energy surface

Both MM and QM methods, with different level of theory, are able to calculate the energy of a molecule and thus can be used to produce the potential energy surface (PES). PES is a hyper-surface of 3N-6 dimensional (where N is the number of atoms of the system) where every point represents the potential energy associated to a certain conformation of a molecule. Along this surface there are some points, called stationary points, that are interesting to describe the energetic of the molecule. Such points are the minimum, maximum and saddle points:

- *Maximum points* are associated with unfavorable conformation of the molecule
- *Saddle points* are associated with transition state geometry that represents the maximum energy point along the reaction coordinate while being a minimum with respect to all other directions on the PES
- *Minimum points* are associated with energetically stable conformations. There are two kinds of minima in a PES, absolute minimum represents the most stable conformation of the molecule whereas local minima are all the other stable geometries that lie in valleys surrounded by energy peaks.

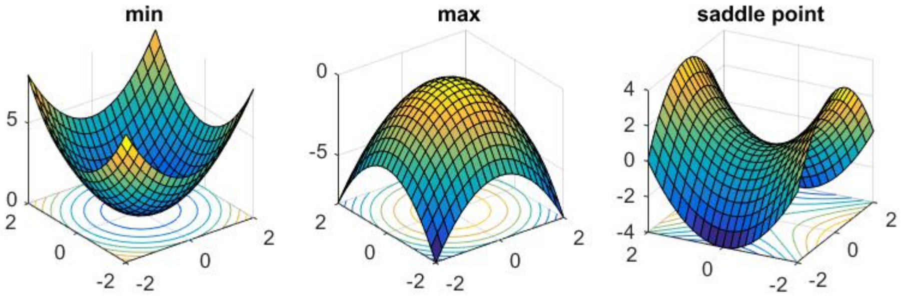


Figure 1.2: Graphical representation of minimum, maximum and saddle point.

Both MM and QM methods can deal with the study of energy minima that represent the populated conformations of a molecular structure. Energy minima can be found through a process referred to as energy minimization. Different types of minimization procedures are available, the most used are:

- *Steepest descent method* → The starting point is an arbitrary point on the PES and the direction of the minimization is opposite to the direction in which the gradient is the largest. After a minimum has been reached, a new minimization is performed along the steepest descent direction from that point. This process is repeatedly carried out since a minimum has been reached in all directions.

- *Conjugate gradient* → it is quite similar to the previous one but to avoid an oscillatory behavior during the optimization, it implements the evaluation of precedent direction to correct the gradient. This strategy guarantees a faster convergence than the steepest descent.
- *Newton-Raphson* → it is based on the Taylor series expansion of the potential energy at a current conformation: the geometry of the molecule is then corrected by taking into consideration both the first and the second derivatives of the series, namely gradient and curvature respectively, of the potential energy. This method is much faster than both steepest descent and conjugate gradient methods but it is computationally rather expensive.

Bibliography

[1] Margareta R. A. Blomberg* And and Per E. M. Siegbahn. A Quantum Chemical Approach to the Study of Reaction Mechanisms of Redox-Active Metalloenzymes. 2001.

[2] Peter Y. Chou. Prediction of Protein Structural Classes from Amino Acid Compositions. In *Prediction of Protein Structure and the Principles of Protein Conformation*, pages 549–586. Springer US, Boston, MA, 1989.

[3] Christopher J Cramer. *Essentials of Computational Chemistry.* Wiley, second edi edition, 2004.

[4] D R Hartree. Mathematical Proceedings of the Cambridge Philosophical Society The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(24):111–132, 1928.

[5] Warren J Hehre. A Guide to Molecular Mechanics and Quantum Chemical Calculations.

[6] Frank Jensen. Introduction to Computational Chemistry.

[7] Martin Karplus and J. Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9(9):646–652, sep 2002.

[8] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, 140(4A):A1133–A1138, nov 1965.

[9] Andrew R. Leach. *Molecular modelling : principles and applications.* Prentice Hall, 2001.

[10] Adam Liwo, editor. *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes*, volume 1 of *Springer Series in Bio-/Neuroinformatics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[11] Jaroaw Meller. Molecular Dynamics.

[12] C. C. J. Roothaan. New Developments in Molecular Orbital Theory. *Reviews of Modern Physics*, 23(2):69–89, apr 1951.

[13] J. C. Slater. Atomic Shielding Constants. *Physical Review*, 36(1):57–64, jul 1930.

[14] Stephen Wilson. Chemistry by Computer. In *Chemistry by Computer*, pages 7–17. Springer US, Boston, MA, 1986.

Part I

Quantum chemical study of the mechanism of the Mo-nitrogenase: characterization of structural and electronic properties of the activate state E_4

Chapter 2

Abstract

Nitrogenase is the metalloenzyme that catalyzes biological N_2 fixation accounting for the major input of fixed nitrogen into the biogeochemical N cycle. This process entails a multielectron/protons reduction by the sequential addition of two electrons and two protons (at each step) to molecular nitrogen resulting in the formation of diazene, hydrazine at the end ammonia. Nitrogen reduction occurs in eight steps: during the first four of which, the activation of the cofactor is obtained by the income of four electrons and four protons; during the last four steps the actual reduction of nitrogen to ammonia occurs along with the production of one molecule of di-hydrogen. This work was focused on the structural characterization of the key cofactor state E_4, that is now accepted as the intermediate that actually bind N_2 in the active site. For this purpose, DFT simulations have been performed on the FeMo-co active site considering all the iron oxidation state assignments proposed in literature ([$5Fe^{III}$:$2Fe^{II}$], [$3Fe^{III}$:$4Fe^{II}$]and [$1Fe^{III}$:$6Fe^{II}$]) with protonated and not protonated forms of homocitrate. Moreover the electronic properties of E_4 has been studied in order to evaluate the spin alignment of ferrous and ferric sites. In this work DFT has been applied to provide a comprehensive evaluation of the available experimental information about the geometric parameters and oxidation states of FeMo-cofactor. Results suggest to rule out [$5Fe^{III}$:$2Fe^{II}$] and [$1Fe^{III}$:$6Fe^{II}$] oxidation states because in this case the optimized geometries are not consistent with EPR/ENDOR data. Only in the cases of [$3Fe^{III}$:$4Fe^{II}$] with a protonated form of homocitrate, the optimized structure of FeMo-co has two bridging hydrides with a common vertex on Fe_6 in according with experimental data. In particular, both hydrides are bridged to the same iron atoms (Fe_6 and Fe_2) and the [Fe-H-Fe] planes are close to 90° as predicted by EPR spectroscopy. BS-DFT simulations reveal an antiferromagnetically coupled spins ($4\downarrow$:$3\uparrow$) on iron atoms, in detail the two Fe atoms that bind hydrides show a parallel spin-coupling pattern ($\downarrow Fe_2$:$\downarrow Fe_6$)

Chapter 3

Introduction

Nitrogen is an essential element contained in many biomolecules necessary for life like proteins and nucleic acids [32, 72]. Nitrogen is abundantly available on Earth as dinitrogen gas N_2 that represent the 79% of the atmosphere, but despite its abundance most living organisms are unable to use nitrogen in this form. In fact N_2 is a highly inert molecule that is stable both under kinetic and thermodynamic point of view [58, 45]. Nitrogen must be fixed to ammonia NH_3or nitrate NO_3^- [21, 23, 75] to be bio-accessible, nitrogenases catalyze the former of these reactions reducing N_2 to NH_3. There are three families of nitrogenase: the "conventional" Mo-nitrogenase [48, 46, 22] are present in every nitrogen-fixing microorganism and it represents the best studied and characterized member of this metalloenzyme; the "unconventional" V-nitrogenase and Fe-nitrogenase are found only in a limited set of N_2-fixing bacteria and are present secondarily to (Mo)-nitrogenase, being expressed only when Mo concentrations are limiting [30, 49]. The Mo-nitrogenase is a two-component metalloenzyme [18, 60] composed of the Fe-protein and MoFe-protein.

Fe-protein, the so-called dinitrogenase reductase or component II, is a homodimeric protein having two ATP-binding sites, one on each subunit, and a single Fe_4-S_4 cluster bridging the two identical subunits [82, 35, 26, 40, 47]. The MoFe-protein, the so-called dinitrogenase or component I, is an α_2 β_2 heterotetramer having two types of metal clusters: the P-cluster Fe_8-S_7 is located at each interface $\alpha\beta$ in the MoFe-protein near the Fe-protein face [26, 63, 46, 57, 55], the FeMo-cofactor, one on each subunit α, which provides the active site for substrate binding and reduction [19, 70]. The accepted mechanism of action of nitrogenase was developed by extensive studies in 70s and 80s by Lowe

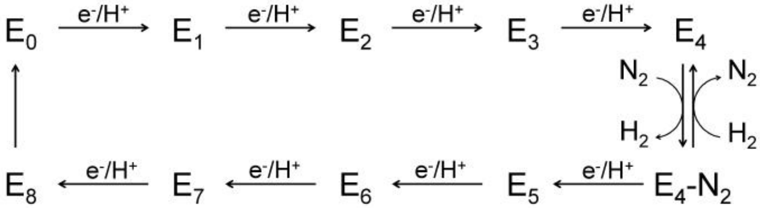


Figure 3.1: Simplified kinetic scheme that highlights correlated electron/proton delivery in eight steps.

and Thorneley and their co-workers [20, 76, 81]. The Lowe-Thorneley (LT) kinetic model describes the N_2 fixation among eight catalytic intermediates denoted E_n, where n is the number of electron/protons delivery to MoFe protein (Figure 3.1).

During each step Fe-protein and MoFe-protein associate and dissociate involving the single electron transfer from the environment to the active site (FeMo-cofactor) and MgATP hydrolysis [35, 94]. According to LT model, the N_2 reduction to NH_3 is given by the following equation [71, 39]:

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

This equation shows two key aspects of the nitrogenase mechanism: the involvement of ATP hydrolysis in the substrate reduction and the formation of one mole of H_2 per mole of N_2 reduced with the apparent waste of two reducing equivalents and four ATP molecules [20]. Hoffmann *et al* describe this obligatory formation of H_2 as functional for N_2 binding and activation. In fact, they proposed the occurrence of a concerted substitution of two hydrides, bound to the same Fe with dinitrogen, following a reductive elimination (RE) process [39]. The RE mechanism implies that the 4 electrons/protons present on the FeMo-co during E_4 are formally two protons and two hydrides. When the two hydrides leave the cofactor as H_2, the Fe atom that binds N_2 becomes highly activated through formal reduction by two equivalents (the release of H_2 carries away only two of the four reducing equivalents stored in E_4). This activates Fe that becomes able to carry out the initial reduction of N_2 transferring the two reducing equivalents to the N_2 (Figure 3.2). In detail, Hoffman and co-worker have suggested to consider the LT model as divided into two stages: the first one, that involves the first four steps, represents the activation of the FeMo-cofactor with the accumulation of four protons/electrons; the second one is the genuine fixation of N_2 in which protons/electrons are added in alternating way to the two nitrogen atoms of N_2 leading to the formation of diazene, hydrazine and finally two NH_3 [20, 76, 77, 25]. E_4, is called "Janus intermediate" because it comes at the halfway point along the entire reaction pathway thus both looking back to the starting point and concomitantly to the end of the fixation process. Due to its peculiar position in the turnover cycle it is thought to be useful to answer the question "why H_2 is lost upon N_2 binding" [39]. Little is known about this catalytic intermediate and in general about the structure and electronic properties of the FeMo-cofactor. FeMo-co has been identified as a [Mo-Fe_7-S_9-X] cluster where X is the interstitial atom present at the middle of the iron cage [68, 43, 42, 44], only recently it has been discovered that the interstitial atom is a C atom [73, 52, 53, 80, 79]. This metal cluster has a homocitrate ligand bonded via the hydroxyl and carboxylate oxygen atoms and it is held with only two residues by the subunit α of the Mo-protein: His_{442} bonded to the molybdenum atom and Cys_{275} bonded to the iron atom located oppositely to Mo [56]. Up to now the protonation state of homocitrate is unknown, however a computational work already suggested the possibility that the coordinated hydroxyl oxygen is protonated in the resting state [64]. Also the oxidation state assignment of Fe and Mo atoms for FeMo-co is still under debate: presently in literature three possible oxidation state are discussed that are associated with the total charge on the cofactor: $[MoFe7S9C]^{-3}$, $[MoFe7S9C]^{-1}$ and $[MoFe7S9C]^{+1}$. All these oxidation states assignments are consistent with a total spin of S=3/2 derived from EPR spectroscopy [29, 83]. In according to the total charge and to the oxidation state of Mo, the oxidation states of iron atoms are reported in the following table[36, 17, 16]:

COFACTOR CHARGE	Mo^{IV}	Mo^{III}
$[MoFe_7S_9C]^{+1}$	$5Fe^{III}$:$2Fe^{II}$	$6Fe^{III}$:$1Fe^{II}$
$[MoFe_7S_9C]^{-1}$	$3Fe^{III}$:$4Fe^{II}$	$4Fe^{III}$:$3Fe^{II}$
$[MoFe_7S_9C]^{-3}$	$1Fe^{III}$:$6Fe^{II}$	$2Fe^{III}$:$5Fe^{II}$

About E_4, it is known the total spin S=1/2 obtained from EPR signal [41] and the presence of two iron-bridging hydrides [M-H-M'] in orthogonal way from ENDOR spectroscopy analysis [39]. As said before, E_4 intermediate is the key step of the N_2 fixation, thus investigating the structure and the electronic properties of this intermediate state can be useful to better understand how nitrogenases are able to reduce the triple bond of N_2. In this study BLYP/DFT simulations were used to carry out a comprehensive evaluation of the data available about E_4 in order to identify the correct oxidation state assignment of metal atoms, the protonation state of homocitrate ligand and the disposition of protons and hydrides on the FeMo-co.

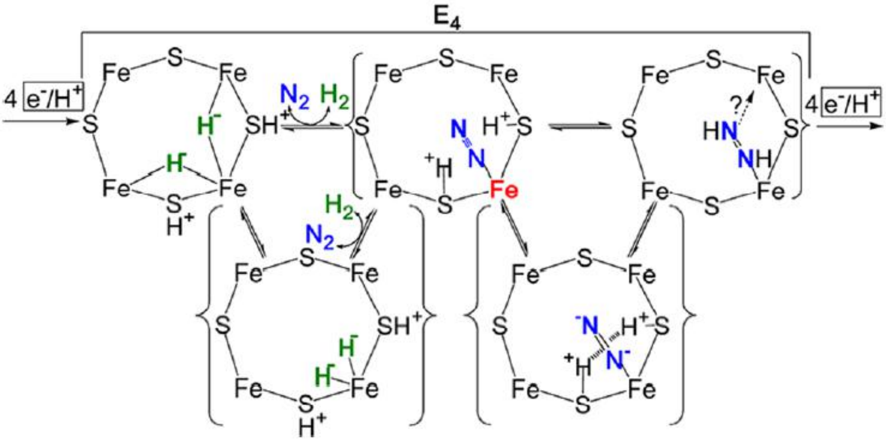


Figure 3.2: Visualization of the reductive elimination mechanism for H_2 release upon N_2 binding to E_4.

Chapter 4

Computational methods

4.1 DFT implementation

DFT calculations have been carried out using the TURBOMOLE 7.0.1 suite of programs [119]. The pure exchange-correlation functional BLYP [147] has been adopted in that it was already validated in literature for QM investigations on nitrogenases [24]. Due to the presence of one unpaired electron, an open-shell system (S=1/2), all simulations were performed using the Unrestricted Hartree–Fock (UHF) that provide the spin α and β electrons with the possibility to occupy spatially distinct orbitals. The computational model of the MoFe protein active site has been obtained starting from X-ray structure 1M1N [95].

4.1.1 Minimal model

A first set of calculations was performed using a minimal model of the active site composed only by FeMo-co and the ligands Cys_{275}, His_{442} and homocitrate truncated to methylthiolate, imidazole and glycolate, respectively. All possible configuration of protons (Figure 4.1) and hydrides (Figure 4.2) were tested considering all possible oxidation state assignments ([$5\mathrm{Fe}^{III}$:$2\mathrm{Fe}^{II}$ / $6\mathrm{Fe}^{III}$:$1\mathrm{Fe}^{II}$], [$3\mathrm{Fe}^{III}$:$4\mathrm{Fe}^{II}$ / $4\mathrm{Fe}^{III}$:$3\mathrm{Fe}^{II}$] and [$1\mathrm{Fe}^{III}$:$6\mathrm{Fe}^{II}$ / $2\mathrm{Fe}^{III}$:$5\mathrm{Fe}^{II}$] that correspond in terms of total charge to $[\mathrm{MoFe7S9C}]^{-3}$, $[\mathrm{MoFe7S9C}]^{-1}$ and $[\mathrm{MoFe7S9C}]^{+1}$, respectively) of metal atoms proposed in literature and both protonated and not protonated form of homocitrate. This means that up to 252 different structures (6 possible protons configurations x 7 possible hydrides configurations x 3 oxidation state assignments x 2 protonated form of homocitrate) have been energetically sampled. All the 252 structures were clustered into six families as reported in the table:

Family	**Oxidation state**	**Homocitrate**	**total charge of system**	**configuration protons/hydrides**
1	[5Fe^{III}:2Fe^{II}]		-2	42
2	[5Fe^{III}:2Fe^{II}]	protonated	-1	42
3	[3Fe^{III}:4Fe^{II}]		-4	42
4	[3Fe^{III}:4Fe^{II}]	protonated	-3	42
5	[1Fe^{III}:6Fe^{II}]		-6	42
6	[1Fe^{III}:6Fe^{II}]	protonated	-5	42

The total charge is calculated considering methylthiolate (-1), imidazole (0), glycolate (-2 or -1 in protonated form) and FeMo-co (sulfur and carbon are taken in their usual closed-shell form S^{-2} and C^{-4}). Due to the small size of the model, the total charge of each structure was neutralized by adding an appropriate number of K^+ atoms. The following mixed basis sets has been used: a triple-ζ basis set with double polarization and diffuse functions for Mo, C interstitial atoms (def2-TZVPD) [158] and S atom (6-311++G**) [37]; a double-ζ basis set (DZP) [59] for all other atoms (ligands). The 252 models were optimized using this protocol:

1. Pre-optimization with keeping hydrides bond distances frozen so as to allow the FeMo-co to adapt itself to the presence of hydrides.
2. Geometry optimization of the entire model without constraints.
3. Single point calculation using COSMO [137] to evaluate the polarization effects of both solvent and amino acids environment (epsilon =40) [117, 104, 105].

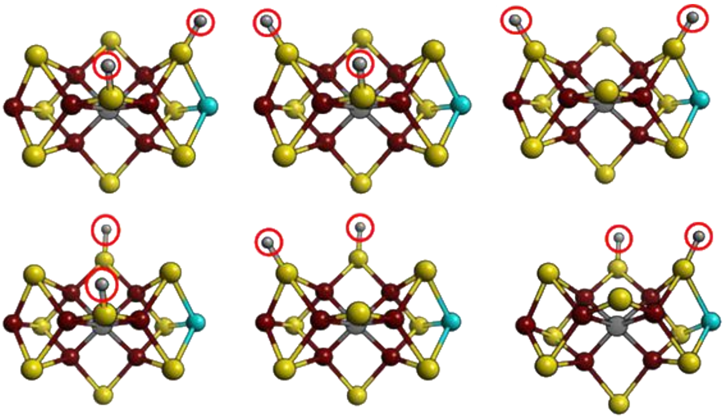


Figure 4.1: Possible configurations of protons on FeMo-co. Protons are highlighted in red.

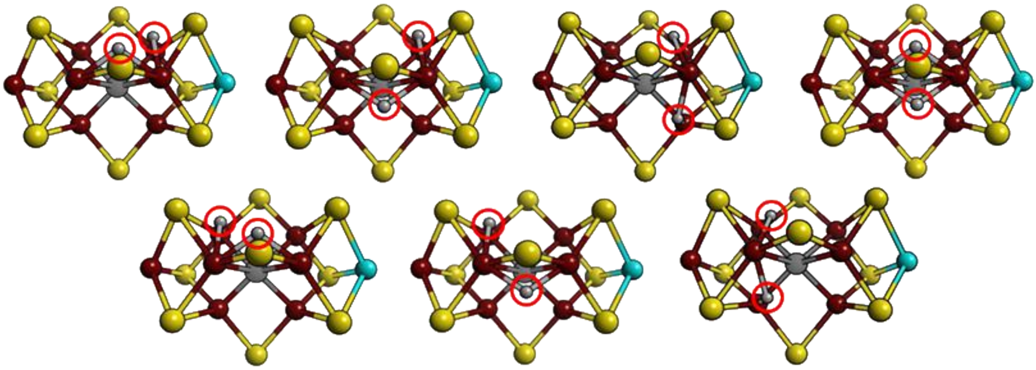


Figure 4.2: Possible configurations of bridging hydrides [Fe-H-Fe] on FeMo-co. Hydrides are highlighted in red.

4.1.2 Enlarged model

All structures from previous optimizations that show a disposition of protons and hydrides in agreement with the experimental data obtained from EPR/ENDOR spectroscopy were used to perform a second set of calculations using a more accurate model to characterize the electronic properties of the active site. For this purpose, the minimal model is enlarged with 10 amino acids near to the active site (in a radius of 4.5 Å): Ser_{278}, His_{195}, Arg_{277}, Phe_{381}, Gln_{191}, Val_{70}, Arg_{96}, Arg_{359}, Ile_{231}, Tyr_{229}. For sake of simplicity and with the aim to preserve their crucial chemico-physical properties, all amino acids have been truncated so as to consider only the functional groups of their side chains. In detail amino acids have been truncated as following:

- Val_{70} and Ile_{231} at C_α carbon
- Tyr_{229}, His_{195}, Ser_{278}, and Phe_{381} at C_β carbon
- Gln_{191} at C_γ carbon
- Arg_{96} and Arg_{359} to C_δ carbon

No counter ions have been used to neutralize the total charge of this model type. Mixed basis sets has been used: a triple-ζ basis set with double polarization and diffuse functions for Mo, Fe, S and C interstitial atoms; a double-ζ basis set (DZP) for all other atoms (amino acids atoms). The optimization procedure follows the same steps used to optimize the minimal models with the introduction of a final step in which different spin alignments were energetically tested by using the broken symmetry (BS) approach [110]. The spin coupling alignments tested for the seven iron sites are based on previous work of Noodleman *et al* about FeMo-co cluster [62, 27].

4.2 Broken symmetry approach

Studying the FeMo-co is challenging because this cluster consists of three layers of high spin irons, all antiferromagnetically coupled. This type of spin-coupled systems is unsuitable for the classic (single-determinant) DFT approach. The reason is that such systems are described by wavefunctions that correspond to a linear combination of more determinants and, unfortunately, canonical SCF-DFT Kohn-Sham electron density is associated with a single-determinant wave function. To calculate the energies and properties of FeMo-co is important to describe the weak antiferromagnetic coupling at the same level of theory as strong metal-metal and metal-ligand bonds. This is possible using the broken symmetry approach (BS) developed by Noodleman *et al* [110]. The BS approach consists in the localization of opposite spin in different fragments of the molecule, in order to use a single-determinant representation of the interaction of exchange within it. After the spin localization it is possible to manually change the orientation of each spin site in order to explore all possible configurations. The number of possible M permutations of spin states in N spin sites are equal to $M = 2^N/2$.

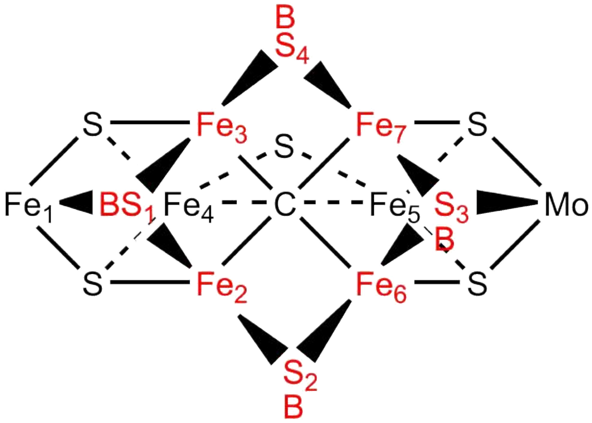


Figure 4.3: Nomenclature scheme of each atom of the FeMo-co. Atoms marked in red represent the catalytic face of the cluster.

Chapter 5

Results and discussion

The first part of the project consisted in identifying the correct configuration of protons and hydrides on the FeMo-co by using the EPR data available in literature as constraints. H ENDOR spectroscopic analysis of E_4 revealed the presence of two hyperfine-coupled metal-bridging hydrides [M-H-M'] with a large isotropic hyperfine coupling (a_{iso}≈24 MHz) and an anisotropic hyperfine contribution very close to a complete rhombicity (T=[13.3, 0.7, 12.7] MHz). Moreover Mo ENDOR measurements showed a very small hyperfine coupling $a_{iso}(^{95}Mo)$ ∼4 MHz. All these data suggest the presence of two hydrides that are bound to Fe atoms in a bridging configuration (Fe-H-Fe). Furthermore the analysis of relative orientation of the hyperfine coupling tensors suggests that the planes of two hydrides are reciprocally orthogonal [39]. Starting from that, all possible configurations of hydrides and protons were tested considering all the six FeMo-co "families" (see computational methods for details). By these simulations, it is possible to rule out the family1 and family2 ([$5Fe^{III}$:$2Fe^{II}$] and [$5Fe^{III}$:$2Fe^{II}$ H^+]) because in these conditions the FeMo-co is not able to bind hydrides (Figure 5.1 panel **a** and **b**). In fact, during the optimizations of both family1 669and family2 members, the FeMo-co undergoes a structural rearrangement that leads to the release of H_2. The difference in energy between the structure that loses H_2 (the most stable structure) and the most stable structure that features two hydrides is more than 20 kcal/mol, that prevents the formation of E_4 intermediate (as it is expected). The other families ([$3Fe^{III}$:$4Fe^{II}$], [$3Fe^{III}$:$4Fe^{II}$ H^+], [$1Fe^{III}$:$6Fe^{II}$ H^+]) allow the formation of structures with two hydrides bridged to the same iron atoms (Fe_2 and Fe_6) (Figure 5.1 panel **c**, **d** and **f**) except for the model with

[1Fe^{III}:6Fe^{II}] oxidation state (Figure 5.1 panel **e**). In the family [1Fe^{III}:6Fe^{II}] the most stable structure shows a terminal hydride on Fe$_6$ in contrast with EPR data. However, the second most stable structure possesses the two iron-bridging hydrides and it is higher in energy only by 6.6 kcal/mol than the lowest energy one (Figure 5.1 panel **e**). Due to the small energy difference between the first and the second structure, also the configuration [1Fe^{III}:6Fe^{II}] cannot in principle be excluded as a possible oxidation state of FeMo-co. The angle between the Fe-H-Fe planes is close to 90° in agreement with hyperfine tensor analysis (Figure 5.2). About the proton possible orientation, all families(3-6) feature a proton on the belt-sulfur atom located on Fe$_6$ and Fe$_2$, with the presence of this proton leading to the formation of a sulfhydryl group bound to only one Fe (Figure 5.1 panel **c**, **d, e** and **f**).

The four energy accessible families ([3Fe^{III}:4Fe^{II}], [3Fe^{III}:4Fe^{II} H$^+$], [1Fe^{III}:6Fe^{II}] and [1Fe^{III}:6Fe^{II} H$^+$]) were studied in greater detail to provide a narrower distribution of possible oxidation states and also to pinpoint the protonation state of homocitrate. For this purpose, the computational model was enlarged with 10 amino acids of the first sphere of coordination (see computational methods). In this second set of calculations, broken symmetry approach has been implemented in order to study all possible spin coupling alignments of iron atoms, furthermore the variation of redox state of iron ions with the resting state E$_0$ has been evaluated. For each of the four all structures residing within an energy range of 10 kcal/mol have been re-optimized. Results suggest that the only possible combination of oxidation state and protonation state of the homocitrate is represented by family4 ([3Fe^{III}:4Fe^{II} H$^+$]) (Figure 5.3). In fact the most stable structure of FeMo-co has the two hydrides bridged to the Fe$_6$ and Fe$_2$ with a conformation in agreement with EPR data. Moreover, this structure shows a build-up of electron density on Fe$_6$ in agreement with the reductive elimination process proposed by Hoffman [39, 38] (Figure 5.3). Fe$_6$ is located under the the Val^{70} that acts as “gatekeeper” controlling the access of the substrate to the active site [115, 65]. For these reason Fe$_6$ is inferred to be involved in substrates binding. The fact that the most reduced iron atom, so the most activated Fe, is Fe$_6$ is a further point showing agreement with experimental suggestions. The broken symmetry approach reveals that the most appropriate spin alignment is the one with 3↑:4↓ with the Fe$_6$ and Fe$_2$ (the hydride bound irons) coupled in a ferromagnetic way (↓Fe$_6$:↓Fe$_2$) (Figure 5.3). In all other systems FeMo-co loses a SH$^-$ group from the catalytic site (Figure 5.3). It is worth noting that a S group release is observed only during the CO inhibition as pointed out by Spatzal *et al* [74]. Due to this observation, this kind of structure can be associated with an inactivated form of the FeMo-co, thereby it is possible to rule out the configuration [3Fe^{III}:4Fe^{II}], [1Fe^{III}:6Fe^{II}] and [1Fe^{III}:6Fe^{II} H^{+}].

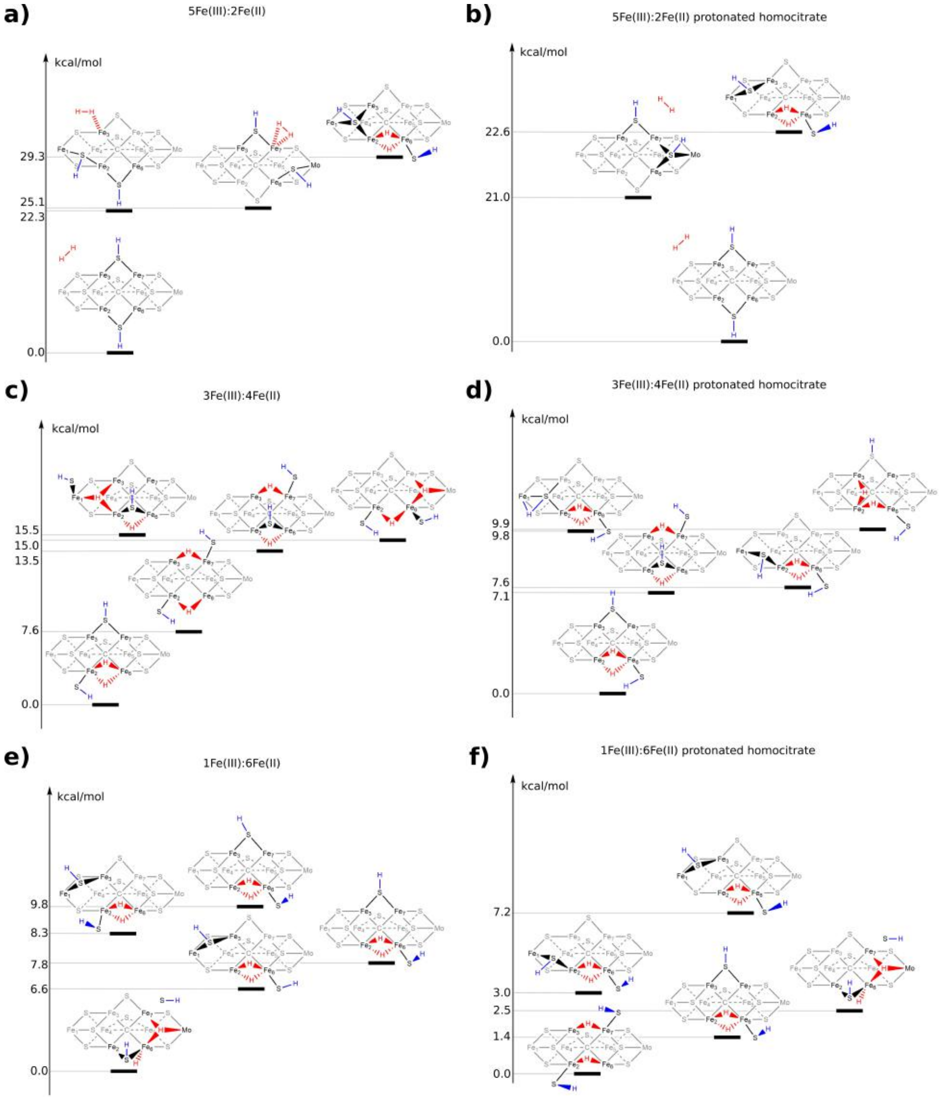


Figure 5.1: Optimization results of models [5Fe^{III}:2Fe^{II}] **(a)**, [5Fe^{III}:2Fe^{II} H^{+}] **(b)**, [3Fe^{III}:4Fe^{II}] **(c)**, [3Fe^{III}:4Fe^{II} H^{+}] **(d),** [1Fe^{III}:6Fe^{II}] **(e)**, [1Fe^{III}:6Fe^{II}H^{+}] **(f).** Hydrides are reported in red and protons in blue.

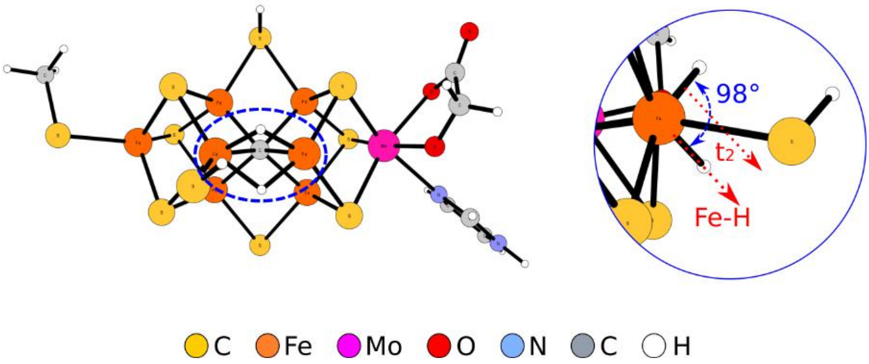


Figure 5.2: Angle between the Fe-H-Fe plane measured for [3Fe^{III}:4Fe^{II}] and [1Fe^{III}:6Fe^{II}] with protonated and non protonated homocitrate. EPR data show that t$_2$$\sim$ 0 for one hydride is parallel to one of the large couplings for the other hydride ($|t_1| \sim |t_2| = 0$) and vice versa, in other word the normal to the [Fe-H-Fe] plane (the t$_2$ direction) of one hydride lies roughly along one of the two Fe-H bond of the other. Angle is reported in blue, t2 and Fe-H plane in red.

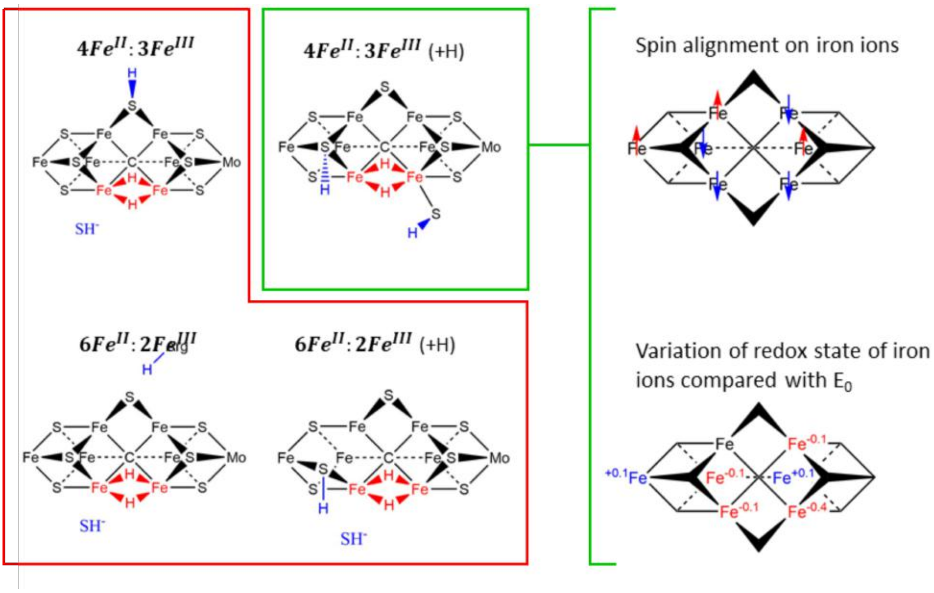


Figure 5.3: Optimization results of enlarged models [3Fe^{III}:4Fe^{II}] , [3Fe^{III}:4Fe^{II} H$^+$] , [1Fe^{III}:6Fe^{II}] and [1Fe^{III}:6Fe^{II}H$^+$]. For model [3Fe^{III}:4Fe^{II} H$^+$] it is reported the best spin alignment and the variation of redox state of iron ions.

Chapter 6

Conclusion

In this contribution a systematic evaluation has been performed of all experimental data available in literature about the FeMo-co: this was a key prerequisite to find out the energetically accessible structures of the E_4 state, representing the most elusive (probably due to its large instability) among the E_n states in the nitrogenase turnover cycle. DFT results suggest that the correct oxidation state of the FeMo-co is [$3Fe^{III}$:$4Fe^{II}$] and also that homocitrate ligand is protonated on the hydroxyl oxygen. This last finding seems reasonable in consideration of the highly negative charge accumulated on the FeMo-co, especially at the E_4 stage. It is proposed a structure of E_4 in good agreement with the experimental data. In particular E_4 shows two iron-bridging hydrides on the Fe_2 and Fe_6 atoms with two proton located on the S_2B and S_1B. Fe_6 is the most reduced iron of the complex and therefore is activated to reduce nitrogen in agreement with the reductive elimination process of H_2. The spin coupling pattern found consists of $3\uparrow$:$4\downarrow$ with a ferromagnetic coupling alignment of Fe_2 and Fe_6.

Bibliography

[15] R. Ahlrichs, M. Bar, M. Haser, H. Horn, and C. Kolmel. Electronic structure calculations on workstation computers: The program system turbomole. *Chemical Physics Letters*, 162(3):165–169, oct 1989.

[16] Ragnar Bjornsson, Frederico A. Lima, Thomas Spatzal, Thomas Weyhermüller, Pieter Glatzel, Eckhard Bill, Oliver Einsle, Frank Neese, and Serena DeBeer. Identification of a spin-coupled Mo(III) in the nitrogenase iron-molybdenum cofactor. *Chemical Science*, 5(8):3096–3103, 2014.

[17] Ragnar Bjornsson, Frank Neese, Richard R. Schrock, Oliver Einsle, and Serena Debeer. The discovery of Mo(III) in FeMoco: Reuniting enzyme and model chemistry. *Journal of Biological Inorganic Chemistry*, 20(2):447–460, 2015.

[18] W A Bulen and J R LeComte. The nitrogenase system from Azotobacter: two-enzyme requirement for N2 reduction, ATP-dependent H2 evolution, and ATP hydrolysis. *Proceedings of the National Academy of Sciences of the United States of America*, 56(3):979–86, sep 1966.

[19] Barbara K. Burgess. The iron-molybdenum cofactor of nitrogenase. *Chemical Reviews*, 90(8):1377–1406, dec 1990.

[20] Barbara K. Burgess and David J. Lowe. Mechanism of Molybdenum Nitrogenase. *Chemical Reviews*, 96(7):2983–3012, 1996.

[21] Donald E. Canfield, Alexander N. Glazer, and Paul G. Falkowski. The Evolution and Future of Earth's Nitrogen Cycle. *Science*, 330(6001):192–196, oct 2010.

[22] M K Chan, J Kim, and D C Rees. The nitrogenase FeMo-cofactor and P-cluster pair: 2.2 A resolution structures. *Science (New York, N.Y.)*, 260(5109):792–4, may 1993.

[23] Qi Cheng. Perspectives in Biological Nitrogen Fixation Research. *Journal of Integrative Plant Biology*, 50(7):786–798, jul 2008.

[24] Ian Dance. The hydrogen chemistry of the FeMo-co active site of nitrogenase. *Journal of the American Chemical Society*, 127(31):10925–10942, 2005.

[25] L C Davis. Hydrazine as a substrate and inhibitor of Azotobacter vinelandii nitrogenase. *Archives of biochemistry and biophysics*, 204(1):270–6, oct 1980.

[26] D R Dean, J T Bolin, and L Zheng. Nitrogenase metalloclusters: structures, organization, and synthesis. *Journal of bacteriology*, 175(21):6737–44, nov 1993.

[27] Lukoyanov Dmitriy, Vladimir Pelmenschikov, Maeser Nathan, Mikhail Laryukhin, Yang Tran Chin, Louis Noodleman, Dennis R. Dean, David A. Case, Lance C. Seefeldt, and Brian M. Hoffman. Testing if the Interstitial Atom, X, of the Nitrogenase Molybdenum-Iron Cofactor Is N or C: ENDOR, ESEEM, and DFT Studies of the S = 3/2 Resting State in Multiple Environments. *Inorg. Chem.*, 46:11437–49, 2007.

[28] S. Duval, K. Danyal, S. Shaw, A. K. Lytle, D. R. Dean, B. M. Hoffman, E. Antony, and L. C. Seefeldt. Electron transfer precedes ATP hydrolysis during nitrogenase catalysis. *Proceedings of the National Academy of Sciences*, 110(41):16414–16419, oct 2013.

[29] Munck E., Rhodes H., Ormejohnson W.H., Davis L.C., W J Brill, and V K Shah. Nitrogenase. VIII. Mossbauer and EPR spectroscopy. The MoFe protein component from Azotobacter vinelandii OP. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 400(1):32–53, jul 1975.

[30] Robert R. Eady. Structure-Function Relationships of Alternative Nitrogenases. *Chem. Rev.*, 96:3013–3030, 1996.

[31] O. Einsle, F Akif Tezcan, Susana L A Andrade, Benedikt Schmid, Mika Yoshida, James B Howard, and Douglas C Rees. Nitrogenase MoFe-Protein at 1.16 A Reso-

lution: A Central Ligand in the FeMo-Cofactor. *Science*, 297(5587):1696–1700, sep 2002.

[32] S J Ferguson. Nitrogen cycle enzymology. *Current opinion in chemical biology*, 2(2):182–93, apr 1998.

[33] M M Georgiadis, H Komiya, P Chakrabarti, D Woo, J J Kornuc, and D C Rees. Crystallographic structure of the nitrogenase iron protein from Azotobacter vinelandii. *Science (New York, N.Y.)*, 257(5077):1653–9, sep 1992.

[34] Stefan Grimme. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry*, 25(12):1463–1473, sep 2004.

[35] R V Hageman and R H Burris. Nitrogenase and nitrogenase reductase associate and dissociate with each catalytic cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 75(6):2699–702, jun 1978.

[36] Travis V. Harris and Robert K. Szilagyi. Comparative assessment of the composition and charge state of nitrogenase FeMo-cofactor. *Inorganic Chemistry*, 50(11):4811–4824, 2011.

[37] Warren J. Hehre. *Ab initio molecular orbital theory.* Wiley, 1986.

[38] Brian M. Hoffman, Dmitriy Lukoyanov, Dennis R. Dean, and Lance C. Seefeldt. Nitrogenase: A draft mechanism. *Accounts of Chemical Research*, 46(2):587–595, 2013.

[39] Brian M. Hoffman, Dmitriy Lukoyanov, Zhi Yong Yang, Dennis R. Dean, and Lance C. Seefeldt. Mechanism of nitrogen fixation by nitrogenase: The next stage. *Chemical Reviews*, 114(8):4041–4062, 2014.

[40] J B Howard and D C Rees. Nitrogenase: A Nucleotide-Dependent Molecular Switch. *Annual Review of Biochemistry*, 63(1):235–264, jun 1994.

[41] Robert Y. Igarashi, Mikhail Laryukhin, Patricia C. Dos Santos, Hong-In Lee, Dennis R. Dean, Lance C. Seefeldt, and Brian M. Hoffman. Trapping H- Bound to the Nitrogenase FeMo-Cofactor Active Site during H2 Evolution: Characterization by ENDOR Spectroscopy. *J. Am. Chem. Soc.*, 127:6231–6241, 2005.

[42] Se Bok Jang, Mi Suk Jeong, Lance C. Seefeldt, and John W. Peters. Structural and biochemical implications of single amino acid substitutions in the nucleotide-dependent switch regions of the nitrogenase Fe protein from Azotobacter vinelandii. *JBIC Journal of Biological Inorganic Chemistry*, 9(8):1028–1033, dec 2004.

[43] Jang S.B, L.C. Seefeldt, and J.W. Peters. Insights into Nucleotide Signal Transduction in Nitrogenase: Structure of an Iron Protein with MgADP Bound. *Biochemistry*, 39:14745–52, 2000.

[44] Mi Suk Jeong and Se Bok Jang. Structural basis for the changes in redox potential in the nitrogenase Phe135Trp Fe protein with MgADP Bound. *Molecules and cells*, 18(3):374–82, dec 2004.

[45] Hong-Peng Jia and Elsje Alessandra Quadrelli. Mechanistic aspects of dinitrogen cleavage and hydrogenation to produce ammonia in catalysis and organometallic chemistry: relevance of metal hydride bonds and dihydrogen. *Chemical Society reviews*, 43(2):547–64, jan 2014.

[46] J Kim and D C Rees. Structural models for the metal centers in the nitrogenase molybdenum-iron protein. *Science*, 257(5077):1677–82, sep 1992.

[47] Jongsun Kim and Douglas C. Rees. Nitrogenase and biological nitrogen fixation. *Biochemistry*, 33(2):389–397, jan 1994.

[48] Jongsun Kim, D. Woo, and D. C. Rees. X-ray crystal structure of the nitrogenase molybdenum-iron protein from Clostridium pasteurianum at 3.0-.ANG. resolution. *Biochemistry*, 32(28):7104–7115, jul 1993.

[49] Jongsun Kirn and D. C. Rees. Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from Azotobacter vinelandii. *Nature*, 360(6404):553–560, dec 1992.

[50] Andreas Klamt. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *The Journal of Physical Chemistry*, 99(7):2224–2235, feb 1995.

[51] Andreas Klamt. Calculation of UV/Vis Spectra in Solution. *The Journal of Physical Chemistry*, 100(9):3349–3353, jan 1996.

[52] K. M. Lancaster, M. Roemelt, P. Ettenhuber, Y. Hu, M. W. Ribbe, F. Neese, U. Bergmann, and S. DeBeer. X-ray Emission Spectroscopy Evidences a Central Carbon in the Nitrogenase Iron-Molybdenum Cofactor. *Science*, 334(6058):974–977, nov 2011.

[53] Kyle M. Lancaster, Yilin Hu, Uwe Bergmann, Markus W. Ribbe, and Serena DeBeer. X-ray Spectroscopic Observation of an Interstitial Carbide in NifEN-Bound FeMoco Precursor. *Journal of the American Chemical Society*, 135(2):610–612, jan 2013.

[54] Lee, Yang, and Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review. B, Condensed matter*, 37(2):785–789, jan 1988.

[55] D J Lowe, K Fisher, and R N Thorneley. Klebsiella pneumoniae nitrogenase: pre-steady-state absorbance changes show that redox changes occur in the MoFe protein that depend on substrate and component protein ratio; a role for P-centres in reducing dinitrogen? *The Biochemical journal*, 292 (Pt 1:93–8, may 1993.

[56] Dmitriy Lukoyanov, Zhi Yong Yang, Nimesh Khadka, Dennis R. Dean, Lance C. Seefeldt, and Brian M. Hoffman. Identification of a Key Catalytic Intermediate Demonstrates That Nitrogenase Is Activated by the Reversible Exchange of N2 for H2. *Journal of the American Chemical Society*, 137(10):3610–3615, 2015.

[57] L Ma, M A Brosius, and B K Burgess. Construction of a form of the MoFe protein of nitrogenase that accepts electrons from the Fe protein but does not reduce substrate. *The Journal of biological chemistry*, 271(18):10528–32, may 1996.

[58] Bruce A. MacKay and Fryzuk Michael D. Dinitrogen Coordination Chemistry: On the Biomimetic Borderlands. *Cgem. Rev.*, 104:385–402, 2004.

[59] L. S. C. Martins, F. E. Jorge, M. L. Franco, and I. B. Ferreira. All-electron Gaussian basis sets of double zeta quality for the actinides. *The Journal of Chemical Physics*, 145(24):244113, dec 2016.

[60] L E Mortenson. Components of cell-free extracts of Clostridium pasteurianum required for ATP-dependent H2 evolution from dithionite and for N2 fixation. *Biochimica et biophysica acta*, 127(1):18–25, sep 1966.

[61] Louis. Noodleman, David A. Case, and Arie. Aizman. Broken symmetry analysis of spin coupling in iron-sulfur clusters. *Journal of the American Chemical Society*, 110(4):1001–1005, feb 1988.

[62] Vladimir Pelmenschikov, David A. Case, and Louis Noodleman. Ligand-bound S = 1/2 FeMo-cofactor of nitrogenase: Hyperfine interaction analysis and implication for the central ligand X identity. *Inorganic Chemistry*, 47(14):6162–6172, 2008.

[63] J W Peters, K Fisher, W E Newton, and D R Dean. Involvement of the P cluster in intramolecular electron transfer within the nitrogenase MoFe protein. *The Journal of biological chemistry*, 270(45):27007–13, nov 1995.

[64] Szilagyi R.K., Musaev D.G., and Morokuma K. Theoretical studies of biological nitrogen fixation. Part II. Hydrogen bonded networks as possible reactant and product channels. *Journal of Molecular Structure: THEOCHEM*, 506(1-3):131–146, jul 2000.

[65] Patricia C. Dos Santos, Robert Y. Igarashi, Hong-In Lee, Brian M. Hoffman, Lance C. Seefeldt, and Dennis R. Dean. Substrate Interactions with the Nitrogenase Active Site. *Acc. Chem. Res.*, 38:208–214, 2005.

[66] Ranjana Sarma, Brett M. Barney, Stephen Keable, Dennis R. Dean, Lance C. Seefeldt, and John W. Peters. Insights into substrate binding at FeMo-cofactor in nitrogenase from the structure of an ??-70Ile MoFe protein variant. *Journal of Inorganic Biochemistry*, 104(4):385–389, 2010.

[67] Ansgar Schäfer, Christian Huber, and Reinhart Ahlrichs. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *The Journal of Chemical Physics*, 100(8):5829–5835, 1994.

[68] Se Bok Jang, Seefeldt Lance C., and Peters John W. Modulating the Midpoint Potential of the [4Fe-4S] Cluster of the Nitrogenase Fe Protein. *Biochemistry*, 39:641–648, 2000.

[69] Lance C. Seefeldt, Brian M. Hoffman, and Dennis R. Dean. Mechanism of Mo-Dependent Nitrogenase. *Annual Review of Biochemistry*, 78(1):701–722, jun 2009.

[70] V K Shah and W J Brill. Isolation of an iron-molybdenum cofactor from nitrogenase. *Proceedings of the National Academy of Sciences of the United States of America*, 74(8):3249–53, aug 1977.

[71] F B Simpson and R H Burris. A nitrogen pressure of 50 atmospheres does not prevent evolution of hydrogen by nitrogenase. *Science (New York, N.Y.)*, 224(4653):1095–7, jun 1984.

[72] Vaclav. Smil. *Enriching the earth : Fritz Haber, Carl Bosch, and the transformation of world food production*. MIT Press, 2001.

[73] T. Spatzal, M. Aksoyoglu, L. Zhang, S. L. A. Andrade, E. Schleicher, S. Weber, D. C. Rees, and O. Einsle. Evidence for Interstitial Carbon in Nitrogenase FeMo Cofactor. *Science*, 334(6058):940–940, nov 2011.

[74] Thomas Spatzal, Perez K.A., Oliver Einsle, James B. Howard, and Douglas C Rees. Ligand binding to the FeMo-cofactor: Structures of CO-bound and reactivated nitrogenase. *Science*, 345(6204):1620–1623, 2014.

[75] Bo Thamdrup. New Pathways and Processes in the Global Nitrogen Cycle. *Annual Review of Ecology, Evolution, and Systematics*, 43(1):407–428, dec 2012.

[76] R N Thorneley and D J Lowe. The mechanism of Klebsiella pneumoniae nitrogenase action. Pre-steady-state kinetics of an enzyme-bound intermediate in N2 reduction and of NH3 formation. *The Biochemical journal*, 224(3):887–94, dec 1984.

[77] ROGER N. F. THORNELEY, ROBERT R. EADY, and DAVID J. LOWE. Biological nitrogen fixation by way of an enzyme-bound dinitrogen-hydride intermediate. *Nature*, 272(5653):557–558, apr 1978.

[78] Spyridon Vicatos, Maite Roca, and Arieh Warshel. Effective approach for calculations of absolute stability of proteins using focused dielectric constants. *Proteins: Structure, Function, and Bioinformatics*, 77(3):670–684, nov 2009.

[79] J. A. Wiig, Y. Hu, C. C. Lee, and M. W. Ribbe. Radical SAM-Dependent Carbon Insertion into the Nitrogenase M-Cluster. *Science*, 337(6102):1672–1675, sep 2012.

[80] Jared A. Wiig, Chi Chung Lee, Yilin Hu, and Markus W. Ribbe. Tracing the Interstitial Carbide of the Nitrogenase Cofactor during Substrate Turnover. *Journal of the American Chemical Society*, 135(13):4982–4983, apr 2013.

[81] P E Wilson, A C Nyborg, and G D Watt. Duplication and extension of the Thorneley and Lowe kinetic model for Klebsiella pneumoniae nitrogenase catalysis using a MATHEMATICA software platform. *Biophysical chemistry*, 91(3):281–304, jul 2001.

[82] H C Winter and R H Burris. Nitrogenase. *Annual Review of Biochemistry*, 45(1):409–426, jun 1976.

[83] K Zimmermann, W.H. Ormejohnson, E. Munck, V.K. Shah, and W.J. Brill. Nitrogenase X: Mossbauer and EPR studies on reversibly oxidized MoFe protein from Azotobacter vinelandii OP Nature of the iron centers. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 537(2):185–207, dec 1978.

Part II

Structural characterization of the nitrogenase MoFe protein with the substrate acetylene trapped near the active site

Chapter 7

Abstract

The biological reduction of dinitrogen to ammonia is catalyzed by the complex metalloenzyme nitrogenase. Structures of the nitrogenase component proteins, Fe protein, MoFe protein, and the stabilized complexes of these component proteins, have been determined, providing a foundation for a number of fundamental aspects of the complicated catalytic mechanism. The reduction of dinitrogen to ammonia is a complex process that involves the binding of N_2 followed by reduction with multiple electrons and protons. Electron transfer into nitrogenase is typically constrained to the unique electron donor, the Fe protein. These constraints have prevented structural characterization of the active site with bound substrate. Recently it has been realized that selected amino acid substitutions in the environment of the active site metal cluster (FeMo-co) allow substrates to persist even in the resting state. Reported here is a 1.70 Å crystal structure of a nitrogenase MoFe protein α-$96^{Arg\rightarrow Gln}$ variant with the alternative substrate acetylene trapped in a channel in close proximity to FeMo-co. Complementary theoretical calculations support the validity of the acetylene interaction at this site and is also consistent with more favorable interactions in the variant MoFe protein compared to the native MoFe protein. This work represents the first structural evidence of a substrate trapped in the nitrogenase MoFe protein and is consistent with earlier assignments of proposed substrate pathways and substrate binding sites deduced from biochemical, spectroscopic, and theoretical studies.

Chapter 8

Introduction

Nitrogenase, which catalyzes the reduction of atmospheric dinitrogen (N_2) to ammonia (NH_3) in bacteria and archaea [88], is an integral component of the global nitrogen cycle [89]. The most abundant nitrogenase, the molybdenum-dependent enzyme, contains two metalloprotein components, MoFe protein and Fe protein [89, 115]. The MoFe protein is an α2β2 heterotetramer that contains 2 [8Fe-7S] P-clusters and 2 active site [7Fe-9S-C-Mo-homocitrate] FeMo-co, and must undergo multiple cycles of MgATP-dependent association and dissociation events to deliver electrons from the Fe protein to the substrate binding and catalytic sites in the MoFe protein [118, 94]. The Fe protein is a homodimer bridged by a [4Fe-4S] cubane and has two sites for MgATP binding and hydrolysis, one in each subunit [96]. The proposed substrate binding site of nitrogenase is located in the MoFe protein at the FeMo-co [88, 106]. Several amino acid sequence comparisons [87] and site-directed mutagenesis studies [101, 85] have shown that certain amino acid residues are important for substrate interactions at the FeMo-co including α-70^{Val}, α-195^{His} and α-96^{Arg} [85, 112, 116, 108]. It has been shown that the substitution of α-70^{Val}, by amino acids with small side chain, resulted in the ability of the MoFe protein from *Azotobacter vinelandii* to reduce larger organic substrates such as propyne, or 1-butyne, which are normally poor substrates for nitrogenase. On the other hand, the substitution of the α-70^{Val} by longer side chain, for example isoleucine, decrease the ability of the MoFe protein reduce a number of substrates including dinitrogen and acetylene, while maintaining normal proton reduction activity, indicating an important role of the α-70^{Val} at the substrate binding site at the MoFe protein [112]. The structural characterization of nitrogenase with bound substrate has been a challenging problem due to the complex interplay between the component proteins and the corresponding high turnover rate of the substrates. The binding of substrates (e.g., acetylene) to the FeMo-co was detected only under turnover conditions where all components of the nitrogenase complex and a reducing agent are present [99]. Previous biochemical and EPR spectroscopy studies have demonstrated the ability of several site directed variants to trap acetylene and cyanide substrates at the active site of MoFe protein under turnover conditions [116, 108] and in the resting state [85, 90]. Acetylene-derived species bound to the FeMo-co in α-$195^{His \rightarrow Gln}$ MoFe protein variant had been characterized under turnover conditions using Q-band EPR and ENDOR spectroscopy [116, 108]. It has been shown that the EPR signals of the dithionite-reduced MoFe protein variants where α-96^{Arg} were substituted by leucine, glutamine, alanine, or histidine are similar to that from native MoFe protein. The spectroscopic characterization of these MoFe protein variants incubated with acetylene or cyanide resulted in changes of the EPR signal, showing interaction of the acetylene and cyanide with FeMo-co in the resting state of the MoFe proteins [85]. To capitalize on this finding, the α-$96^{Arg \rightarrow Gln}$ variant of MoFe protein was crystallized in the presence of acetylene. Reported here are the first crystal structure and computational analysis of the nitrogenase MoFe protein variant α-$96^{Arg \rightarrow Gln}$ with acetylene captured near the FeMo-co active site.

Chapter 9

Computational methods

9.1 DFT

DFT calculations have been carried out using the TURBOMOLE 7.0.1 suite of programs [119]. The following mixed basis sets has been used:

- a split valence triple-ζ basis set with polarization (TZVP) [158] for FeMo-co (E_0, S=3/2 state), α-442^{His}, α-275^{Cys}, homocitrate (that are all covalently bound to Fe-S cluster) and acetylene;
- a triple-ζ basis set with double polarization and diffuse functions (def2-TZVPD) [113] for Mo atom;
- a single-ζ basis set (SVP) [111] for amino acid side chain atoms.

The pure exchange-correlation functional BLYP [147] has been adopted in that it was already validated in literature for QM nitrogenase investigations [92]. Dispersion correction to SCF energy has been performed using the approach proposed by Grimme for DFT [137]. Polarization effects of both solvent and active site amino acid side chains have been evaluated according to the conductor-like screening model (COSMO, $\varepsilon = 40$) [117, 104, 105]. Spin configurations of iron atoms within FeMo-co (determined by a Broken Symmetry approach) [110] have been verified to correspond to that generally accepted for E_0 i.e., all iron ions are reciprocally antiferromagnetically coupled, in line with previous investigations [109]. DFT geometry optimization included the FeMo-co/acetylene complex and relevant amino-acid residues (α-65^{Ala}, α-69^{Gly}, α-70^{Val}, α-96^{Gln}, α-191^{Gln}, α-195^{His}, α-229^{Tyr}, α-231^{Ile}, α-275^{Cys}, α-278^{Ser}, α-359^{Arg}, α-381^{Phe}, α-442^{His}) starting from crystal structure of α-96$^{Arg\rightarrow Gln}$ MoFe variant. Amino acid residues have been truncated at Cα so all the atoms of sidechains are present in the model. To our knowledge, this is the largest size setup related to QM-only (i.e., no atom in the model is treated by semi-empiric and/or classic Force Field based methods) investigations on nitrogenase. The original purpose of maintaining zwitterionic forms of amino acids has been subsequently dropped as it yields unrealistic bond dissociations during simulations. Twelve molecules of water identified by crystallization have been included in calculation to simulate explicitly the presence of solvent. Both possible protonation states of homocitrate covalently bound to Fe-S cluster have been considered. This system is studied considering two charge state assignments: the first state has a total charge equal to -4 according to the FeMo-co and amino acids present in the model; the other one is a neutral charge model. The effect of neutralizing the -4 charge of the computational model that we have set up in the present study has been investigated by placing 4 K^+ ions at the exact position where it is located the centroids of the ammonium and the three guanidinium groups belonging to α-304^{Lys}, α-28^{Arg}, α-203^{Arg}, and α-439^{Arg} in the X-ray structure.

The computational model of native MoFe protein active site has been obtained starting from X-ray structure 1M1N [95]; the amino acids residues taken into explicit account were exactly the same used for the α-96$^{Arg\rightarrow Gln}$ MoFe variant structure, with the (trivial) sole exception of α-96^{Arg}. Two sets of calculations have been performed for native MoFe protein: one is with total charge equal to -3 (arising from FeMo-co and amino acids present in the DFT model of NATIVE system); the second is with total charge 0 (-3 charge has been neutralized by three K^+, placed according to criteria previously described). The same corrections to pure DFT energy that were used for the α-96$^{Arg\rightarrow Gln}$ MoFe variant protein have also been applied to the native MoFe protein model (dispersion correction, implicit solvation model and counterpoise correction). Different rotamers of Arg^{96} have been energetically sampled [127].

9.2 Refinement levels used in DFT calculation

A first level refinement has been performed including effects of dispersion forces in the model. This was necessary to treat the attractive component of the potential energy curve associated with London-type interactions [137, 138], which is normally missing in standard density functionals, due to the fact that the stabilizing effect of dispersion occurs from medium- to long-range distance (where electron densities based on truncated basis sets do not overlap).

When using the dispersion correction, the total energy is given by

$$E_{DFT-Disp} = EKS - DFT + E_{disp}$$

where E_{KS-DFT} is the usual self-consistent Kohn-Sham energy as obtained from the chosen functional and E_{dips} is an empirical dispersion correction given by

$$Edisp = -s_6 \sum_{i=1}^{N_{at}-1} \sum_{j=i+1}^{N_{at}} \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij})$$

Here, N_{at} is the number of atoms in the system, C_{ij}^6 denotes the dispersion coefficient for atom pair ij, s_6 is a global scaling factor that only depends on the DF used and R_{ij} is an interatomic distance. The interatomic C_{ij}^6 term is calculated as geometric mean of the form

$$C_6^{ij} = \sqrt{C_6^i C_6^j}$$

$f_{dmp}$is a damping function used to avoid near-singularities for small R.

A second level refinement has required single-point corrections including effect of polarizing environment of the active site pocket ($\varepsilon = 40$, as often used for a large number of protein active sites [117]). The solvation effect is evaluated using COSMO (Conductor-like Screening Model) approach [146]. COSMO is a continuum solvation model (CSM), where the solute molecule forms a cavity within the dielectric continuum of permittivity ε that represents the solvent. The charge distribution of the solute polarizes the dielectric medium. The response of the medium is described by the generation of screening charges on the cavity surface. The screening charges are scaled by a factor

$$f(\varepsilon) = \frac{\varepsilon-1}{\varepsilon+\frac{1}{2}}$$

A third level refinement has been applied evaluating (and correcting by means of Counterpoise method [86]) the so-called BSSE (Basis Set Superposition Error), a computational issue originating in QM calculations (with finite size basis sets, namely all those employed) when evaluating energetics of bi-molecular complex formation (starting from separate molecules). In a DFT simulation, when a monomer A comes close to monomer B, the dimer AB can be artificially stabilized as monomer A uses the extra basis functions from monomer B to describe its electron distribution and vice versa. This is not an error itself, the error arises from the inconsistent treatment of the monomers because they are able to use additional functions from the other monomer only at shorter intermolecular distances and not at large distance. The BSSE approach permits to correct this error by estimating the amount by which monomer A is artificially stabilized by the extra basis functions from monomer B and vice versa.

$$E_{BSSE}(A) = E_A^{AB}(A) - E_A^A(A)$$

$$E_{BSSE}(B) = E_B^{AB}(B) - E_B^B(B)$$

where superscripts are the basis used, the subscripts are the geometry and symbols in brackets are the chemical system considered.

$\Delta E_{binding}$ was calculated using this formula:

$$\Delta E_{binding} = E_{AB} - (E_A + E_B)$$

Where E_{AB}, E_A, E_B and is the energy of FeMo-co/acetylene complex, FeMo-co (all atoms included in "AB" but those of acetylene) and acetylene, respectively. Entropic contributions, necessary to switch from $\Delta E_{binding}$ to $\Delta G_{binding}$, were not explicitly computed due to the very large size (for QM methods) of the system under investigation. The computational cost of the Hessian matrix evaluation (a mandatory task for determining the molecular partition function that allows one to estimate the entropy) is indeed prohibitive for a QM system with ca 270 atoms. Molecular Dynamics simulations on similar protein-small ligand complex suggest, however, that the entropy loss upon binding of substrate within the cavity of the enzyme is approximately 3 kcal/mol, a result that does not significantly alter the values of $\Delta E_{binding}$ [100].

Chapter 10

Results and discussion

This work is in collaboration with research group of Professor John Peters from Montana State University. They have expressed in *Azotobacter vinelandii* a variant of nitrogenase MoFe protein with the substitution $96^{Arg \rightarrow Gln}$. The $96^{Arg \rightarrow Gln}$ MoFe protein was crystallized in an acetylene-saturated environment obtaining a structure with a resolution of 1.7 Å. The analysis of both 2Fo-Fc and Fo-Fc maps reveal an elongated electron density nearby (approximately 4.0 Å) the four Fe atoms of the FeMo-co that are proposed to be involved in substrate interactions [85]. This elongated electron density well fits with the presence of acetylene in the FeMo-co active site.

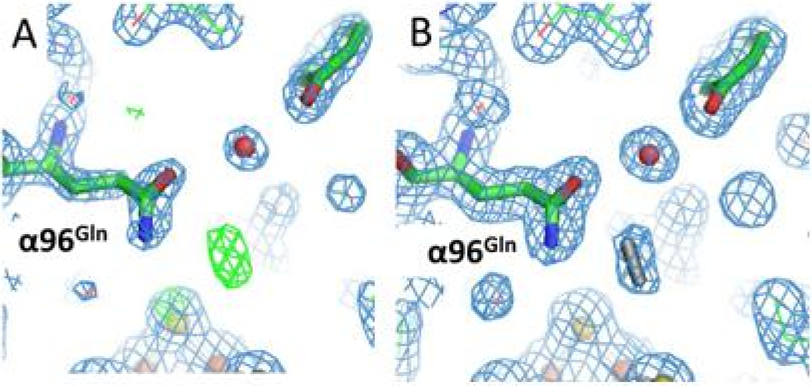


Figure 10.1: Binding of the acetylene at the FeMo-cofactor binding site. Calculated 2Fo-Fc (blue mesh) and Fo-Fc (green mesh) electron density maps without modeling acetylene (A) and modeling acetylene (B) during refinement confirming the presence of acetylene at the FeMo-cofactor binding site.

10.1 Computational analysis of acetylene location near FeMo-co

In order to evaluate the assignment of an acetylene molecule at the position near FeMo-co we have performed a computational analysis. The results of the calculations show that the position of the acetylene in the crystal structure is energetically favorable (Figure 10.2). The energetics of acetylene binding at the various levels of modelling used in the study are summarized in Table 10.1. All interatomic distances (Fe-S, Fe-C, Mo-X bonds) of FeMo-co in the X-ray structure are quite well reproduced by DFT (on average, differing less than 0.06 Å). Figure 10.2 and Figure 10.3 illustrates the location of the acetylene in which it undergoes a partial rotation about one of its C2 axes, which causes it to come closer to FeMo-co than it is located in the crystal structure. The extent of rearrangement looks, however, compatible with both experimental structure at the solved resolution, and with the dispersive nature of the weak interaction between FeMo-co/acetylene. DFT results thus confirm the assignment of the acetylene molecule in close proximity to FeMo-co of the α-$96^{Arg \rightarrow Gln}$ MoFe variant, as observed by X-ray crystallography. A summary of the interatomic distances for the crystal structure and two DFT structures is illustrated in Table 10.2. All data reported in Figure 10.2, Table 10.1, and Figure 10.3 are based on simulations featuring protonated homocitrate (the most likely, in the light of the highly negative charge of FeMo-co). Simulations carried out by using deprotonated homocitrate revealed a very disfavored docking process of acetylene into the nitrogenase active site pocket and have therefore been dropped in subsequent analysis.

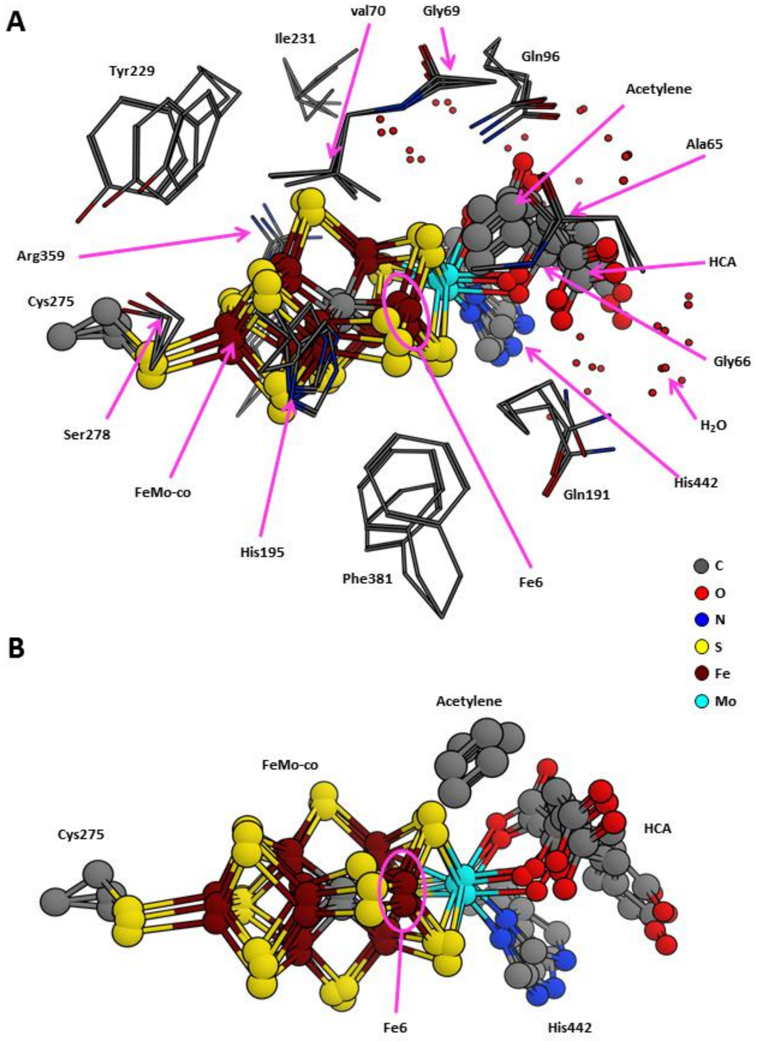


Figure 10.2: Acetylene position at FeMo-co binding site. **(A)** Superposition of FeMo-co and surrounding first-sphere residues of α-96$^{Arg \rightarrow Gln}$ variant crystal structure, DFT optimized structure and DFT optimized structure including dispersion contributions. **(B)** Zoomed-in depiction of FeMo-co and acetylene in the three structures. Color scheme: red-oxygen, blue-nitrogen, grey-carbon, yellow-sulfur, rust red-iron and cyan-molybdenum.

LEVEL OF MODELLING	$\Delta E_{binding}$ (kcal/mol)
DFT/BLYP	2.1
DFT/BLYP/dispersion	-17.3
DFT/BLYP/dispersion COSMO ($\varepsilon = 40$)	-13.0
DFT/BLYP/dispersion COSMO ($\varepsilon = 40$)/BSSE (BSSE_avg)	-4.0 (-8.5)[1]

Table 10.1: Summary of the energetics of binding at the various levels of modelling. The average of the uncorrected and the counterpoise-corrected values is reported in parenthesis (see Ref[1])

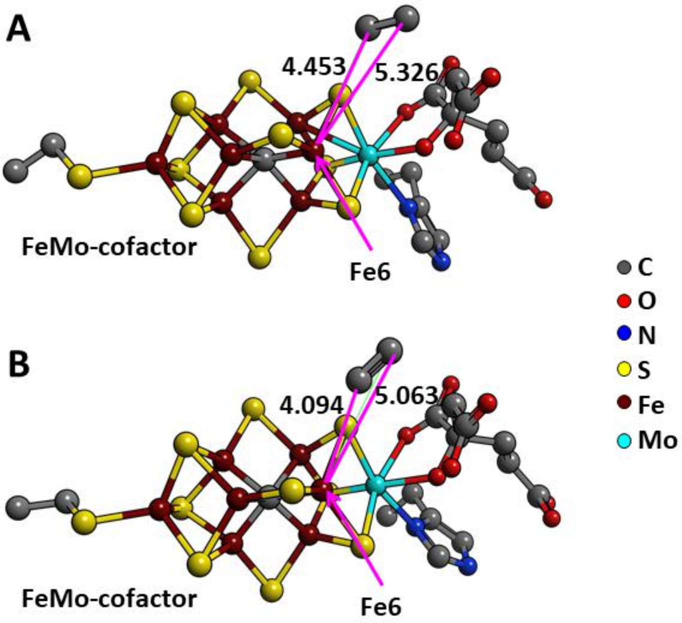


Figure 10.3: Comparison of the acetylene location at the FeMo-cofactor. The acetylene location at the FeMo-cofactor is compared between X-ray **(A)** and DFT optimized structures **(B)**. The interatomic distances are shown in red lines.

	X-Ray	**Mod. cha. -4 DFT + disp**	**Variance Mod. -4 Vs. X-Ray**	**Mod. cha. 0 DFT + disp**	**Variance Mod. 0 Vs. X-Ray**
S$_{Cys275}$**-Fe**$_1$	2.385	2.327	-0.058	2.273	-0.112
N$_{His442}$**-Mo**	2.425	2.339	-0.086	2.300	-0.125
O1$_{HCA}$**-Mo**	2.112	2.186	0.074	2.091	-0.021
O2$_{HCA}$**-Mo**	2.203	2.199	-0.004	2.124	-0.079
C$_{int}$**-Fe**$_6$	1.944	2.077	0.133	1.980	0.036
C$_{int}$**-Fe**$_7$	1.956	1.978	0.022	1.980	0.024
C$_{int}$**-Fe**$_3$	1.970	1.936	-0.034	2.096	0.126
C$_{int}$**-Fe**$_2$	1.962	1.974	0.012	1.929	-0.033
C$_{int}$**-Fe**$_4$	1.964	1.927	-0.037	1.945	-0.019
C$_{int}$**-Fe**$_5$	1.933	1.991	0.058	1.991	0.058
Fe1-S$_1$	2.125	2.267	0.142	2.278	0.153
Mo-S$_2$	2.320	2.421	0.101	2.427	0.107
Fe$_6$**-C1**$_{ETD}$	4.453	4.094	-0.359	4.162	-0.291
Fe$_6$**-C2**$_{ETD}$	5.326	5.063	-0.263	5.120	-0.206

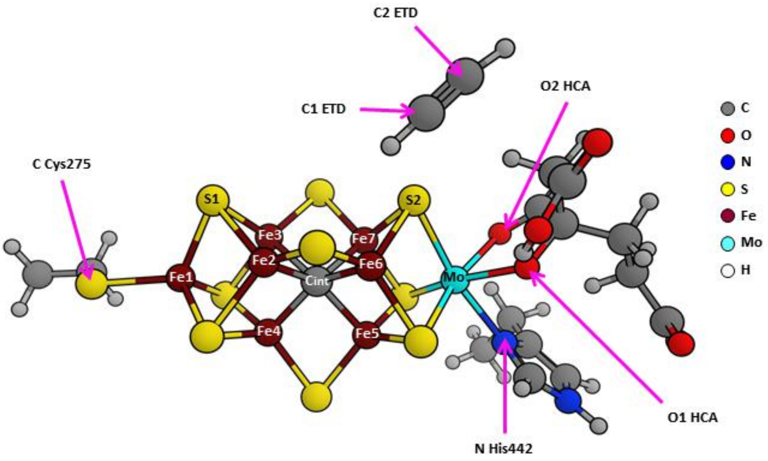


Table 10.2: Main interatomic distances (in Å) for X-Ray structure and the two DFT optimized models, bearing different total charge.

10.2 DFT study of acetylene binding in proximity of a neutrally charged model of α-96$^{Arg \to Gln}$ FeMo-co

The effect of neutralizing the -4 charge of the computational model set up in the present study has been investigated by placing four K^+ ions at the exact position where the crystal structure detected the centroid of the ammonium plus three guanidinium groups belonging to α-304^{Lys} and α-28^{Arg}, α-203^{Arg} and α-439^{Arg}, respectively (Figure 10.4, Figure 10.5, and Table 10.3). The simplest level of theory used (DFT/ BLYP) shows that the slight endergonic binding has to be considered in the light of the precision limit of DFT. Therefore $+2.1$ kcal/mol may be actually indicative of a "near-to-equilibrium" situation (i.e., $\Delta E_{binding} \approx 0$). When switching to more sophisticated treatments of the system nature under investigation (thus considering dispersion interaction attractive holes and refining polarization effects by solvent medium), the predicted binding energy is clearly indicative of an exergonic process. The quite large absolute value of the binding energy (though COSMO screening model partially mitigates this effect) seems to be overestimating an interaction that is expected to be weak (due to the induced dipole moments in acetylene). However, when considering the BSSE effect (both entirely and its half also, as reported by some authors [145]) the issue is actually fixed. At this point the only contribution left to include would be the entropy cost paid by acetylene to access the FeMo-co site. Yet, frequency calculations on such a large system have proved to be prohibitive. Qualitative considerations can be made, however, regarding this point: the positive contribution to the free energy variation ($G > 0$) of the accommodation process within the active site pocket, arising from entropy loss is expected to be paid by substrate already in the diffusion channel leading from outside the protein directly to its active site. Thus, when acetylene gets out of the channel to move in proximity of FeMo-co, where it has been detected by crystallographic analysis and DFT, it has practically lost most of its entropy content. The present results are in line with a thermodynamically facile accommodation of acetylene close to FeMo-co of α-96$^{Arg \to Gln}$ MoFe variant active site.

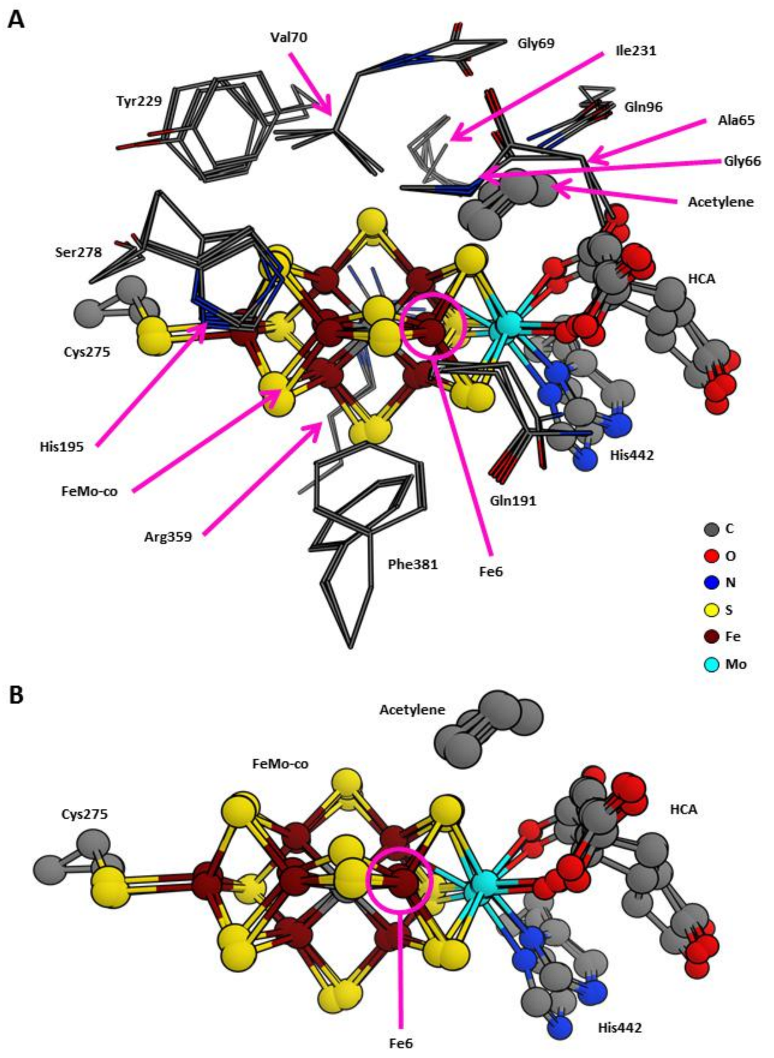


Figure 10.4: Location of the Acetylene at the FeMo-cofactor models with different charge. **(A)** The superposition of FeMo-co and first-shell residues of the α-$96^{Arg\rightarrow Gln}$ variant crystal structure and two DFT optimized models (all including dispersion energy effects). **(B)** Zoomed-depiction of the FeMo-co and acetylene in three superposed structures (XRD/DFT[charge=-4]/DFT[charge=0]). The location of the acetylene in the α-$96^{Arg\rightarrow Gln}$ variant crystal structure is compared to FeMo-co model with total charge -4 and neutral charge. Charge is neutralized by four K^+ ions located at the centroid of positively charged residues. Hydrogen and K^+ ions have been omitted from the figure for the simplicity. Color scheme: red-oxygen, blue-nitrogen, gray-carbon, yellow-sulfur, rust red-iron and cyan-molybdenum.

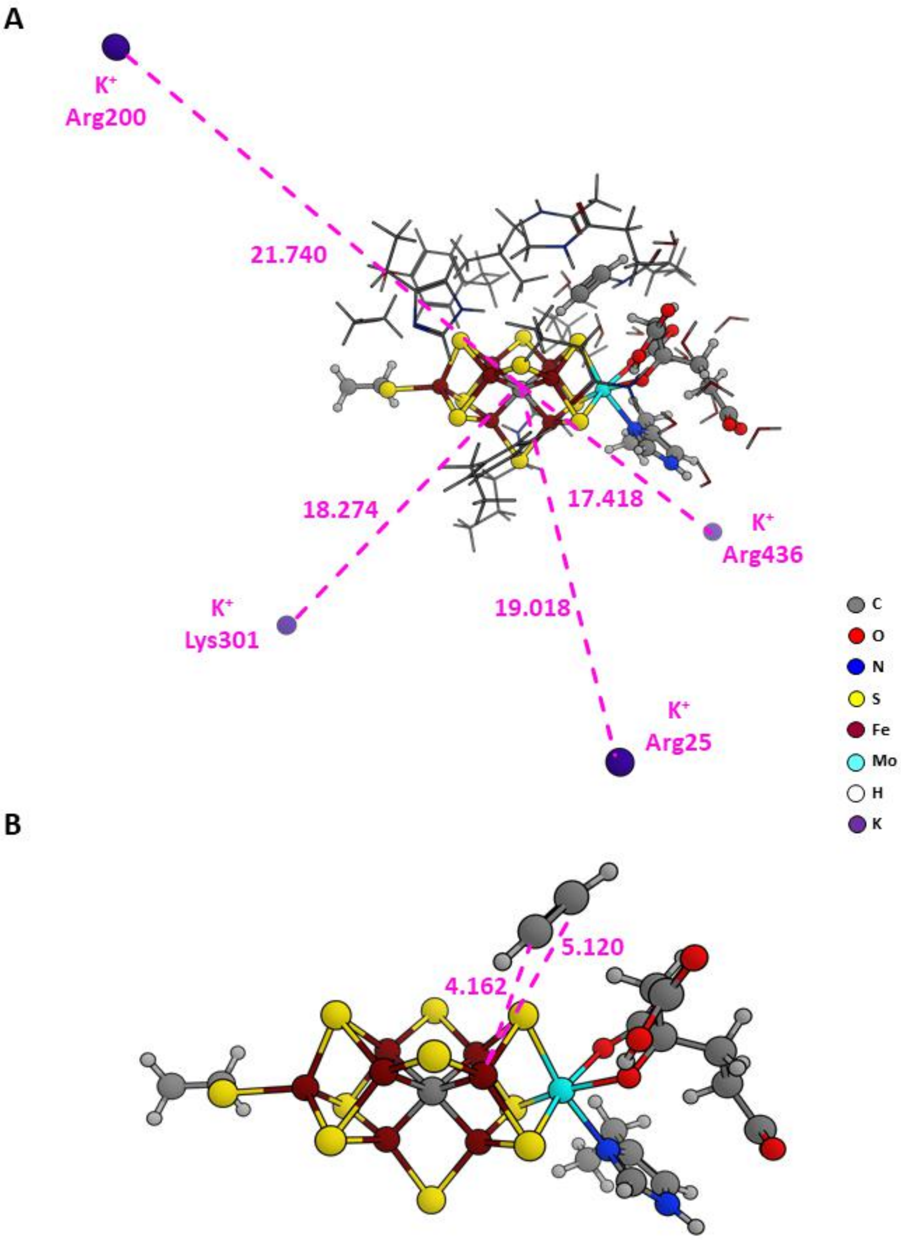


Figure 10.5: Binding of acetylene at the neutrally charged FeMo-co model with four K^+ ions. **(A)** The distances between K^+ ions and interstitial carbon of FeMo-co are displayed in red dashed lines. The amino acid residues modelled by K^+ ions are shown as red spheres. **(B)** The distances between Fe_6 of FeMo-cofactor and acetylene after the DFT converged simulation are shown in red dashed lines.

	Total charge model = -4		**Total charge model = 0**	
	DFT	**DFT/solvation ($\varepsilon = 40$)**	**DFT**	**DFT/solvation ($\varepsilon = 40$)**
ΔE	-17.3	-13.0	-18.1	-17.5
ΔE + BSSE **(BSSE_avg)**[1]	-8.3 (-12.8)[1]	-4.0 (-8.5)[1]	-9.1 (-13.6)[1]	-8.5 (-13.0)[1]

Table 10.3: Comparative summary of binding energy of acetylene to active site, as computed by including different effects (total charge, solvation and counterpoise correction to BSSE; in parenthesis the correction is averaged as described in Ref[1]). Present data already include dispersion effects (that are thus omitted in the shown notation). Energy is reported in kcal/mol.

10.3 Acetylene binding to native MoFe protein active site pocket

Computational studies were also carried out on the native MoFe protein active site pocket to determine if acetylene biding is more or less favorable compared to the α-$96^{Arg \to Gln}$ MoFe variant. Table 10.4 summarizes the energetics of binding at the various levels of modeling used in the study. The results show that the acetylene binding energy to the active site pocket is more favorable for the variant than for the native MoFe protein for each set of calculations. Also, the quantitative difference (α-$96^{Arg \to Gln}$ MoFe protein versus native) in terms of binding energy is compatible with the non-polar nature of the substrate.

An explanation can be found in the different chemico-physical properties (electrostatic features) of the region available for acetylene binding in proximity of FeMo-co: the substitution of Arg to Gln makes the pocket environment (Figure 10.6) more hydrophobic than that of the native MoFe protein. That is clearly expected in the light of the charged-to-neutral residue substitution. Considering the nonpolar nature of acetylene, the α-$96^{Arg \to Gln}$ MoFe variant protein therefore has active site features that make it more suitable for substrate (and analogues) accommodation. It is worth noting that this result is purely thermodynamic and indicates that, taken for granted the kinetic accessibility of acetylene into the region close to FeMo-co both in native and in α-$96^{Arg \to Gln}$ MoFe protein, it would be better accommodated within the α-$96^{Arg \to Gln}$ MoFe catalytic pocket than within that of the native MoFe protein. In addition, it could be considered that the substitution of Arg to Gln increases the size of the binding pocket and along with van der Waals contacts between acetylene and amino acid side of the active site stabilize the position of the acetylene near the FeMo-co.

Table 10.5 reports a comparison of the most relevant interatomic distances for α-$96^{Arg\rightarrow Gln}$ MoFe protein crystal structure and the two DFT optimized models with total charge =0 (variant and native). The scope is to evaluate possible structural variations (movements) of the substrate in α-$96^{Arg\rightarrow Gln}$ variant and native MoFe proteins. One of the two carbon atoms slightly moves in the native (with respect to α-$96^{Arg\rightarrow Gln}$), probably to find better accommodations in response to the Q-to-R switch. EPR measurements of the α-$96^{Arg\rightarrow Gln}$ MoFe variant have demonstrated that the substrate acetylene persists at the resting state of FeMo-co [85]. Other biochemical and spectroscopic studies have suggested the location of substrate reduction to occur at the Fe6 position of the FeMo-co [112, 131]. An earlier spectroscopic study of a MoFe protein variant with α-195^{His} changed to Gln found acetylene bound to FeMo-cofactor. It was proposed from these studies that the acetylene is bound across an Fe-S face of FeMo-co. The studies reported here are on a MoFe protein with a different amino acid substitutions that does not result in a bound acetylene, and thus the two works provide complimentary findings about substrate interaction with nitrogenase.

	ΔE$_{binding}$ (kcal/mol)			
	CHARGE	**DFT/BLYP dispersion**	**DFT/BLYP dispersion COSMO ($\varepsilon = 0$)**	**DFT/BLYP dispersion COSMO ($\varepsilon = 0$) BSSE (BSSE_avg)**
R96Q	-4	-17.3	-13.0	-4.0 (-8.5)[1]
	0	-18.1	-17.5	-8.5 (-13.0)[1]
WT	-4	-16.2	-12.3	-3.3 (-7.8)[1]
	0	-14.8	-13.3	-4.3 (-8.8)[1]

Table 10.4: Summary of the energetics of binding at various levels of modeling used in comparison between the wild type and variant proteins.

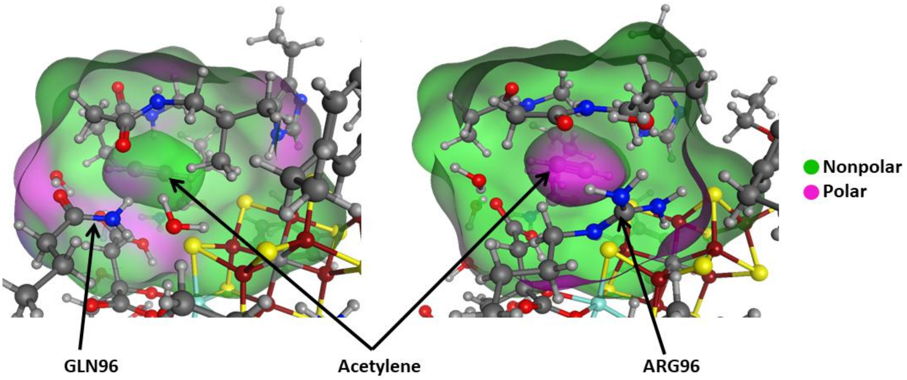


Figure 10.6: Comparison of the electrostatics within the acetylene binding pocket between native (right) and α-96$^{Arg \rightarrow Gln}$ MoFe variant (left) enzymes. Green regions represent nonpolar environment whereas magenta is used for polar environment characterization.

	Å				
Interatomic distance	**X-Ray**	**R96Q**	**Variance X-Ray Vs. R96Q**	**WT**	**Variance X-Ray Vs. WT**
$\mathbf{S}_{Cys275}$**-Fe**$_1$	2.385	2.273	-0.112	2.273	-0.112
$\mathbf{N}_{His442}$**-Mo**	2.425	2.300	-0.125	2.259	-0.166
$\mathbf{O1}_{HCA}$**-Mo**	2.112	2.091	-0.021	2.123	0.011
$\mathbf{O2}_{HCA}$**-Mo**	2.203	2.124	-0.079	2.165	-0.038
$\mathbf{C}_{int}$**-Fe**$_6$	1.944	1.980	0.036	1.978	0.034
$\mathbf{C}_{int}$**-Fe**$_7$	1.956	1.980	0.024	1.989	0.033
$\mathbf{C}_{int}$**-Fe**$_3$	1.970	2.096	0.126	2.125	0.155
$\mathbf{C}_{int}$**-Fe**$_2$	1.962	1.929	-0.033	1.933	-0.029
$\mathbf{C}_{int}$**-Fe**$_4$	1.964	1.945	-0.019	1.936	-0.028
$\mathbf{C}_{int}$**-Fe**$_5$	1.933	1.991	0.058	1.998	0.065
Fe1-S$_1$	2.125	2.278	0.153	2.305	0.180
Mo-S$_2$	2.320	2.427	0.107	2.421	0.101
Fe$_6$**-C1**$_{ETD}$	4.453	4.162	-0.291	3.965	-0.488
Fe$_6$**-C2**$_{ETD}$	5.326	5.120	-0.206	5.052	-0.274

Table 10.5: Comparison of the most relevant interatomic distances (in Å) for R96Q X-ray structure and the two DFT optimized models with total charge = 0 (mutant and wild type). The scope is to evaluate possible structural variations (moves) of the substrate on going from R96Q to WT system. One of the two carbon atoms slightly moves in the WT (respect to R96Q), probably to find better accommodation in response to the Q-to-R switch.

Chapter 11

Conclusions

The α-96$^{Arg\rightarrow Gln}$ MoFe variant crystal structure reveals the position of the substrate acetylene at the FeMo-co. This study provides additional evidence for the possible location and interaction of substrates and inhibitors at the active site of the MoFe protein. The positions and interaction of the α-96^{Gln}, α-70^{Val}, and α-98^{Asn} side chains allows the acetylene molecule to be trapped and energetically stabilized near FeMo-co without the presence of the Fe protein, or MgATP. Rigorous computational calculations show favorable binding of acetylene at the site observed in the crystal structure. Together these findings support that nitrogenase binds substrates at the α-96^{Arg}, α-70^{Val} proximal 4Fe face of the FeMo-co, which is of key mechanistic importance in the reduction of substrates by nitrogenase.

Bibliography

[84] R. Ahlrichs, M. Bar, M. Haser, H. Horn, and C. Kolmel. Electronic structure calculations on workstation computers: The program system turbomole. *Chemical Physics Letters*, 162(3):165–169, oct 1989.

[85] P M Benton, S M Mayer, J Shao, B M Hoffman, D R Dean, and L C Seefeldt. Interaction of acetylene and cyanide with the resting state of nitrogenase alpha-96-substituted MoFe proteins. *Biochemistry*, 40(46):13816–25, nov 2001.

[86] S.F. Boys and F. Bernardi. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Molecular Physics*, 19(4):553–566, oct 1970.

[87] K E Brigle, W E Newton, and D R Dean. Complete nucleotide sequence of the Azotobacter vinelandii nitrogenase structural gene cluster. *Gene*, 37(1-3):37–44, 1985.

[88] B.K. Burgess and D.J. Lowe. Mechanism of Molybdenum Nitrogenase. 1996.

[89] Robert H Burris. THE JOURNAL OF BIOLOGICAL CHEMISTRY Nitrogenases. 266(15):9339–9342, 1991.

[90] J Christiansen, V L Cash, L C Seefeldt, and D R Dean. Isolation and characterization of an acetylene-resistant nitrogenase. *The Journal of biological chemistry*, 275(15):11459–64, apr 2000.

[91] Christopher J Cramer. Essentials of Computational Chemistry Theories and Models Second Edition.

[92] Ian Dance. The Hydrogen Chemistry of the FeMo-co Active Site of Nitrogenase. 2005.

[93] Patricia C Dos Santos, Suzanne M Mayer, Brett M Barney, Lance C Seefeldt, and Dennis R Dean. Alkyne substrate interaction within the nitrogenase MoFe protein. *Journal of inorganic biochemistry*, 101(11-12):1642–8, nov 2007.

[94] Simon Duval, Karamatullah Danyal, Sudipta Shaw, Anna K Lytle, Dennis R Dean, Brian M Hoffman, Edwin Antony, and Lance C Seefeldt. Electron transfer precedes ATP hydrolysis during nitrogenase catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, 110(41):16414–9, oct 2013.

[95] O. Einsle, F Akif Tezcan, Susana L A Andrade, Benedikt Schmid, Mika Yoshida, James B Howard, and Douglas C Rees. Nitrogenase MoFe-Protein at 1.16 A Resolution: A Central Ligand in the FeMo-Cofactor. *Science*, 297(5587):1696–1700, sep 2002.

[96] M M Georgiadis, H Komiya, P Chakrabarti, D Woo, J J Kornuc, and D C Rees. Crystallographic structure of the nitrogenase iron protein from Azotobacter vinelandii. *Science (New York, N.Y.)*, 257(5077):1653–9, sep 1992.

[97] Stefan Grimme. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry*, 25(12):1463–1473, sep 2004.

[98] Stefan Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry*, 27(15):1787–1799, nov 2006.

[99] B D Howes, K Fisher, and D J Lowe. Nitrogenase of <i>Klebsiella pneumoniae</i> : electron nuclear double resonance (ENDOR) studies on the substrate reduction site. *Biochemical Journal*, 297(2):261–264, jan 1994.

[100] Sheeba Jem Irudayam and Richard H. Henchman. Entropic Cost of Protein-Ligand Binding and Its Dependence on the Entropy in Solution. *The Journal of Physical Chemistry B*, 113(17):5871–5884, apr 2009.

[101] H.M. Kent, M. Baines, C. Gormal, B.E. Smith, and M. Buck. Analysis of site-directed mutations in the α- and β-subunits of Klebsiella pneumoniae nitrogenase. *Mol. Microbiol.*, 4:1497–504, 1990.

[102] K S Kim, P Tarakeshwar, and J Y Lee. Molecular Clusters of pi-Systems: Theoretical Studies of Structures, Spectra, and Origin of Interaction Energies. *Chemical reviews*, 100(11):4145–86, nov 2000.

[103] A Klamt and G Schuurmann. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. (2), 1993.

[104] Andreas Klamt. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *The Journal of Physical Chemistry*, 99(7):2224–2235, feb 1995.

[105] Andreas Klamt. Calculation of UV/Vis Spectra in Solution. *The Journal of Physical Chemistry*, 100(9):3349–3353, jan 1996.

[106] David M Lawson and Barry E Smith. Molybdenum nitrogenases: a crystallographic and mechanistic view. *Metal ions in biological systems*, 39:75–119, 2002.

[107] Lee, Yang, and Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review. B, Condensed matter*, 37(2):785–789, jan 1988.

[108] Hong-In Lee, Morten Sorlie, Jason Christiansen, Ruitian Song, Dennis R. Dean, Brian J. Hales, and Brian M. Hoffman. Characterization of an Intermediate in the Reduction of Acetylene by the Nitrogenase α-Gln195 MoFe Protein by Q-band EPR and 13C,1H ENDOR. 2000.

[109] Dmitriy Lukoyanov, Vladimir Pelmenschikov, Nathan Maeser, Mikhail Laryukhin, Tran Chin Yang, Louis Noodleman, Dennis R. Dean, David A. Case, Lance C. Seefeldt, and Brian M. Hoffman. Testing if the Interstitial Atom, X , of the Nitrogenase Molybdenum-Iron Cofactor Is N or C: ENDOR, ESEEM, and DFT Studies of the S = 3\2 Resting State in Multiple Environments. *Inorganic Chemistry*, 46(26):11437–11449, dec 2007.

[110] Louis. Noodleman, David A. Case, and Arie. Aizman. Broken symmetry analysis of spin coupling in iron-sulfur clusters. *Journal of the American Chemical Society*, 110(4):1001–1005, feb 1988.

[111] Dmitrij Rappoport and Filipp Furche. Property-optimized Gaussian basis sets for molecular response calculations. *The Journal of Chemical Physics*, 133(13):134105, oct 2010.

[112] Ranjana Sarma, Brett M Barney, Stephen Keable, Dennis R Dean, Lance C Seefeldt, and John W Peters. Insights into substrate binding at FeMo-cofactor in nitrogenase from the structure of an alpha-70(Ile) MoFe protein variant. *Journal of inorganic biochemistry*, 104(4):385–9, apr 2010.

[113] Ansgar Schafer, Hans Horn, and Reinhart Ahlrichs. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *The Journal of Chemical Physics*, 97(4):2571–2577, aug 1992.

[114] Ansgar Schafer, Christian Huber, and Reinhart Ahlrichs. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *The Journal of Chemical Physics*, 100(8):5829–5835, apr 1994.

[115] Lance C. Seefeldt, Brian M. Hoffman, and Dennis R. Dean. Mechanism of Mo-Dependent Nitrogenase. *Annual Review of Biochemistry*, 78(1):701–722, jun 2009.

[116] M Sorlie, J Christiansen, B J Lemon, J W Peters, D R Dean, and B J Hales. Mechanistic features and structure of the nitrogenase alpha-Gln195 MoFe protein. *Biochemistry*, 40(6):1540–9, feb 2001.

[117] Spyridon Vicatos, Maite Roca, and Arieh Warshel. Effective approach for calculations of absolute stability of proteins using focused dielectric constants. *Proteins: Structure, Function, and Bioinformatics*, 77(3):670–684, nov 2009.

[118] Zhi-Yong Yang, Karamatullah Danyal, and Lance C. Seefeldt. Mechanism of Mo-Dependent Nitrogenase. pages 9–29. 2011.

Part III

Mechanistic investigation on ester hydrolysis catalyzed by *de novo* designed peptide TRIL9CL23H and comparison to carbonic anhydrase.

Chapter 12

Abstract

DFT has been used to investigate the catalytic mechanism of *de novo* design protein TRIL9CL23H, an analogous of carbonic anhydrase (CA) that catalyzes the reversible hydration $CO_2 \leftrightarrow HCO_3^-$. The intimate mechanism proposed for TRIL9CL23H is based on structural features of both α-anhydrase and β-anhydrase: like α-anhydrase also in the de novo designed system a Zn^{2+} ion is coordinated by three histidine residues plus a water molecule, whereas, similarly to β-anhydrase an acid residue (in TRIL9CL23H a glutamate) works as proton acceptor from active site. Unlike carbonic anhydrase, results show that the rate determining step is the release of acid from active site and not the activation of the coordinated water. DFT results indicate that the slight lower catalytic activity of TRIL9CL23H compared to CA can be attributed also to the lack of a residue assisting the final step of acid release. Moreover, a particular mutation (Leu_{19} to Thr) has been investigated through the effect it produces on proton transfer step (Zn^{2+}-H_2O activation). Results suggest this mutation works in the opposite way respect to what originally expected: the activation barrier is indeed higher than in wild type system. Finally, a possible binding site has been proposed for p-nitrophenyl acetate (p-NPA), a substrate of the catalyst. Molecular docking allowed to pinpoint different families of binding sites: two are probably detrimental for catalytic efficiency due to disfavored location, whereas other two (one exterior and one interior) are more likely to be involved in catalysis.

Chapter 13

Introduction

Proteins constitute fundamental macromolecules that cells rely on for a number of tasks: catalysis, recognition, defense, signaling, structural integrity and locomotion. For this reason engineering proteins for improved functions or new applications is an important challenge in medicine and biotechnology field. *De novo* design of proteins allows to build a protein with specific structure and activity, as well as to introduce novel enzymatic activity. There are two possible strategies to design a catalytic protein: the first method consists of inserting a catalytic site into a natural protein; the second method involves the simultaneous design of the entire backbone structure and sequence, the so called ab initio design. Obviously the second approach is a harder task compared to the first one but it has the advantage of providing greater freedom in terms of structures and activities design [153, 130]. However, so far *de novo* design has shown many limitations: first, the number of engineered reactions is few, to the extent that just hydrolytic enzymes are well characterized and designed; another issue is often the low catalytic efficiency of designed proteins and finally, the difficulty is to obtain structures more complex than helical bundles [126, 133, 149]. De novo design is also important in that it allows to improve the knowledge about the folding theory and the relationship between structure and functions of polypeptides [129]. Some of the most studied proteins by *de novo* design are metalloenzymes. These proteins contain particular metal atoms in the active site that catalyze some of the most relevant multi-step chemical processes in nature such as photosynthesis, respiration and water oxidation. Much effort has been devoted toward understanding how metalloproteins work in order to build a novel protein with novel activities. Zinc(II) ion is widely diffused in metalloproteins and in particular the most studied and characterized system is carbonic anhydrase (CA), catalyzing the reversible hydration of carbon dioxide to hydrogen carbonate

$$CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$$

However they can also hydrolyse esters such as p-nitrophenyl acetate (p-NPA) by similar mechanism to the hydration reaction. CA is well-known for its high catalytic activity, in fact it is one of the most efficiently enzyme in nature [161]. For all these reasons CA has been taken as model to build by true *de novo* design a hydrolytic metalloenzyme: TRIL9CL23H, which is a trimer formed by three α-helices in a coiled coil structures (3SCC) based on the scaffolds of TRI family peptides. It has two metal-binding sites: in the first a Hg^{2+} ion is located which provides structural stability, while in the second site there is a Zn^{2+} ion which is important for catalytic activity. Zn^{2+} ion is coordinated by the protein residues forming a pseudo-tetrahedral geometry; in particular it is coordinated by three histidines and one water molecule like in the CA active site [165]. Like CA, TRIL9CL23H can hydrate CO_2 and can also hydrolyze esters such as p-nitrophenyl acetate (p-NPA). Aim of this work has been understanding and characterizing the hydrolytic mechanism associated with p-NPA substrate by using DFT. Specifically, the proposed reaction pathway is similar to that of CA and consists of three steps: first is the H_2O activation to OH^- trough proton release from Zn-bound water to a proton acceptor; the second one is nucleophilic attack of OH^- on the C=O carbon and subsequent ester hydrolysis; the last one is regeneration of the active site. The purpose is to characterize the full catalytic cycle so as to detect possible reasons underlying the slightly lower activity of TRIL9CL23H compared to the natural CA. In a second stage of the work, it has been investigated the effect of replacing Leu_{19} by threonine on the proton transfer associated with Zn-water activation both with DFT method and with ab initio pK_a calculation. Finally, docking simulations have been carried out in order to find (a) possible binding site(s) for p-nitrophenyl acetate to the protein before catalytic events take place.

Chapter 14

Computational methods

14.1 DFT

DFT calculations have been carried out using the TURBOMOLE suite of programs [119], using a split valence triple-ζ basis set with polarization on all atoms (TZVP) [158]. All simulations have been performed using two different exchange-correlation functionals: the hybrid exchange-correlation functional B3LYP[123, 147] that it is known to provide reliable description of systems including transition metals and involving hydrogen-bond interactions [166, 124]; and the pure exchange-correlation functional BP86 [122, 154]. Stationary points of energy hypersurface have been located by means of energy-gradient techniques. Starting atomic positions have been derived from crystal structure (PDB ID 3PBJ). An explicit micro-solvation model has been employed to simulate the presence of water, acting not only as solvent, by playing also an active role in proton shuttling. The optimization of transition-state structures has been performed according to a procedure based on a pseudo-Newton-Raphson method. Initially geometry optimization of transition-state guess geometry is performed keeping the distances corresponding to the selected reaction coordinate constrained. Vibrational analysis (Hessian calculated analytically) at the B3LYP/TZVP and BP86/TZVP level of constrained minimum-energy structures is then carried out, and if one negative eigenmode corresponding to the reaction coordinate is found, the curvature determined at such a point is used as starting point in the transition-state search. The location of the transition-state structure is carried out using an eigenvector-following search: the eigenvectors in the Hessian are sorted in ascending order, with the first one being that associated with the negative eigenvalue. After the first step, however, the search is performed by choosing the critical eigenvector with a maximum overlap criterion, which is based on the dot product with the eigenvector followed in the previous step.

14.2 Refinement levels used in DFT calculation

During all simulations, the effects of dispersion forces have been considered using the D3 dispersion [137, 138] corrections tool of TURBOMOLE. The solvation effect has been taken into account performing a single-point energy correction with COSMO [146] approach ($\varepsilon = 80$). Entropic contributions, necessary to switch from $\Delta E_{binding}$ to $\Delta G_{binding}$, were explicitly computed using the partition function obtained from the full vibrational spectrum analysis. All thermodynamic data have been evaluated for a temperature $T = 298.15\ K$.

14.2.1 Partition function in statistical mechanics

The macroscopic result of an experimental measurement is connected with properties of the microscopic system [142]. The connection between properties of a microscopic system and a macroscopic sample is provided by statistical mechanics. At a temperature of 0 K all molecules are in their energetic ground state but at a finite temperature there is a distribution of molecules in all possible quantum energy states. The relative probability P of a molecule being in a state with an energy ε at a temperature T is given by a Boltzmann factor:

$$P \propto e^{\frac{-\varepsilon}{kT}}$$

The exponential dependence on the energy means that there is a low, but not zero, probability to find a molecule in a high-energy state. This decreased probability for high-energy states is partly offset by the fact that there are many more states with high energy than low. The most probable energy of a molecule in a macroscopic ensemble is therefore not necessarily the one with lowest energy. The partition function (Q) allows to know how many microstates are accessible to the system in a given ensemble, in particular it is the sum over all distinct energy levels multiplied with a degeneracy factor g_n that indicates how many states there are with the same energy (E_n).

$$Q = \sum^{n} g_n e^{\frac{-E_n}{kT}}$$

In case of a molecular system the partition function is the sum of electronic, vibrational, rotational and translational contributions:

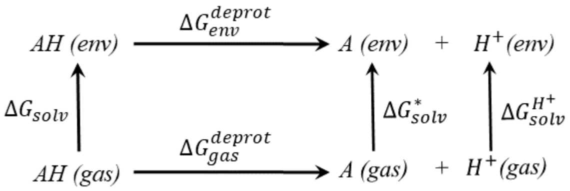
$$Q_{tot} = Q_{elec} + Q_{vib} + Q_{rot} + Q_{trans}$$

14.3 pK$_a$ calculation

The pK_a values in aqueous solution have been evaluated using JAGUAR 8.0 [125] from SCHRODINGER suite of programs. pK_a value can be calculated with the formula

$$pK_a = -logK_a = \frac{\Delta G^{deprot}_{en}}{RTln10}$$

Unfortunately it is not possible to calculated directly the value of deprotonation free energy but it can be computed from a standard thermodynamic cycle show in the following scheme [127].



According to the thermodynamic cycle the deprotonation free energy is equal to

$$\Delta G^{deprot}_{env} = \Delta G^{deprot}_{gas} + \Delta G^{*}_{solv} + \Delta G^{H^+}_{solv} - \Delta G_{solv}$$

In detail, pK_a prediction module requires the following quantum chemical calculations: geometry optimization of the protonated and deprotonated species by DFT/B3LYP/6-31G* [156, 155] method; accurate single-point energy calculation using B3LYP [123, 147] as exchange-correlation functional and cc-pVTZ(+) [132, 144, 163] basis set for atoms involved in deprotonation reaction while cc-pVTZ [132, 144, 163] covers the rest; solvation free energy calculation of protonated and deprotonated species. The starting geometries for pK_a evaluation are obtained by previous DFT full optimizations, therefore all atoms have been kept fixed except those directly involved in deprotonation reaction.

14.4 Molecular docking

Docking simulations have been carried out with GLIDE 5.9 [134, 139], by SCHRODINGER. GLIDE is a grid-based method program and allows a systematic search of the conformational and positional phase space of docked ligands. Both rigid docking and induced fit docking have been performed: in the first case receptor is held rigid; in the second case ligand binding induces substantial conformation changes in the receptor (side chains are optimized by PRIME 3.2 [187, 186]). The docked conformers are evaluated using glide (G)score and a single best pose per ligand is generated as output:

$$GScore = a * vdW + b * Coul + lipo + Hbond + Metal + BuryP + RotB + Site$$

where vdW is van der Waals energy, *Coul* is Coulomb energy, *Lipo* is lipophilic contacts, *Hbond* indicates hydrogen-bonding contributions, *Metal* indicates ligand-metal binding enthalpy, *BuryP* stands for penalty associated with buried polar group, *RtoB* indicates penalty for freezing rotatable bonds, *Site* denotes polar interactions in the active site, a=0.065 and b=0.130.

The glide score is an empirical scoring function that it is an approximation of the ligand binding free energy. Force field OPLS_2005 [169] is used for all docking simulations (Tables 14.1-14.2).

RIGID DOCKING	
Parameter	**Value**
Precision	SP
Ligand Sampling	FLEXIBLE
Sample Nitrogen Inversion	YES
Sample Ring Conformation	YES
Sampling Torsions for Amides	PENALIZED NONPLANAR CONF.
Add Epik State Penalties to Dock Score	YES
Scaling Factor vdW	0.8
Partial Charge Cutoff	0.15
Core Pattern Comparison	NO
Max Number of Poses per Ligand	50

Table 14.1: Parameters used for rigid docking simulations.

INDUCE FIT DOCKING	
Parameter	**Value**
Protocol	IDF_2006
Box Center	Zn_31
Box Size	40Å
Constraints (H-bond/metal)	O_32
Ligands (Sample Ring Conformation)	YES
Ligands (Amide Bonds)	PENALIZED NONPLANAR CONF.
Glide Docking (Preparation)	YES
Glide Docking (Trim Chains)	D:19_LEU | E:19_LEU | F:19_LEU
Prime Refinement	50Å
Glide Re-docking (Precision)	SP

Table 14.2: Parameters used for Induce Fit Docking docking simulations.

14.4.1 Rigid docking

Glide uses a hierarchical series of filters to search for possible locations of the ligand in the active-site region of the receptor. The shape and properties of the receptor are represented on a grid by several different sets of fields that provide progressively more accurate scoring of the ligand poses. Conformational flexibility is handled in Glide by an extensive conformational search, augmented by a heuristic screen that rapidly eliminates unsuitable conformations, such as conformations that have long-range internal hydrogen bonds. each ligand is divided into a core region and some number of rotamer groups. Each rotamer group is attached to the core by a rotatable bond, but does not contain additional rotatable bonds. The core is what remains when each terminus of the ligand is severed at the “last” rotatable bond. Carbon and nitrogen end groups terminated with hydrogen ($—CH_3$, $—NH_2$, $—NH_3^+$) are not considered rotatable because their conformational variation is of little significance. During conformation generation, each core region is represented by a set of core conformations, the number of which depends on the number of rotatable bonds, conformationally labile rings, and asymmetric pyramidal trigonal nitrogen centers in the core. Every rotamer state for each rotamer group attached to the core is enumerated. The core plus all possible rotamer-group conformations is docked as a single object. For each core conformation an exhaustive search of possible locations and orientations is performed over the active site of the protein. The search begins with the selection of site points on an equally spaced 2 Å grid that covers the active site region. Distances from the site point to the receptor surface are evaluated at a series of pre-specified directions and sorted into distance ranges of width 1 Å. Likewise, distances from the ligand center (the midpoint of the two most widely separated atoms) to the ligand surface are sorted into bins of width 1 Å. For a given site point, the distance ranges from the site point to the receptor are compared with those from the ligand center to the ligand surface. Glide positions the ligand center at the site point if there is a good enough match, but skips over the site point if there is not. The second stage of the hierarchy begins by examining the placement of atoms that lie within a specified distance of the line drawn between the most widely separated atoms (the ligand diameter). This is done for a pre-specified selection of possible orientations of the ligand diameter. If there are too many steric clashes with the receptor, the orientation is skipped. Next, rotation about the ligand diameter is considered, and the interactions of a subset consisting of all atoms capable of making hydrogen bonds or ligand-metal interactions with the receptor are scored. If this score is good enough, all interactions with the receptor are scored. The scoring in these three tests is carried out using Schrödinger’s discretized version of the

ChemScore empirical scoring function. This stage is called *greedy scoring*, because the actual score for each atom depends not only on its position relative to the receptor but also on the best possible score it could get by moving ±1 Å in x, y, or z. Only a small number of the best refined poses is passed on to the third stage in the hierarchy/energy minimization on the pre-computed OPLS-AA van der Waals and electrostatic grids for the receptor. The energy minimization typically begins on a set of van der Waals and electrostatic grids that have been "smoothed" to reduce the large energy and gradient terms that result from too-close interatomic contacts. It finishes on the full-scale OPLS-AA nonbonded energy surface. This energy minimization consists only of rigid-body translations and rotations when external conformations are docked. When conformations are generated internally, however, the optimization also includes torsional motion about the core and end-group rotatable bonds. Unless otherwise specified, a small number of the top-ranked poses are then subjected to a sampling procedure in which alternative local minima core and rotamer-group torsion angles are examined to try to improve the energy score. Finally, the minimized poses are re-scored using Schrödinger's proprietary (G)Score scoring function.

14.4.2 Induce fit docking

In standard docking simulations, ligands are docked into the binding site of a receptor where the receptor is held rigid and the ligand is free to move. However, the assumption of a rigid receptor can give misleading results, since in reality many proteins undergo side-chain or backbone movements, or both, upon ligand binding. These changes allow the receptor to alter its binding site so that it more closely conforms to the shape and binding mode of the ligand. This is often referred to as "induced fit" and is one of the main complicating factors in structure-based drug design. The induced fit docking begins by docking the active ligand with GLIDE. In order to generate a diverse ensemble of ligand poses, the procedure uses reduced van der Waals radii and an increased Coulomb-vdW cutoff, and can temporarily remove highly flexible side chains during the docking step. For each pose, a PRIME structure prediction is then used to accommodate the ligand by reorienting nearby side chains. These residues and the ligand are then minimized. Finally, each ligand is re docked into its corresponding low energy protein structures and the resulting complexes are ranked according to a scoring function combining (G)Score and Prime energies.

Chapter 15

Setting up the model

The model-system used has been set up using the pdb file 3PBD, in particular it has been used the trimer with Zn-bound water and not the trimer with Zn-bound Cl^-. The model core includes: Zn^{2+} cation bonded to H_2O and three imidazole rings belonging to the three histidine residues His_{23} of each chain (Figure 15.1). In order to reproduce approximately the restraining effect on the ZnN_3O core by the missing part of the protein environment, histidine β-carbons have been fixed during all optimizations.

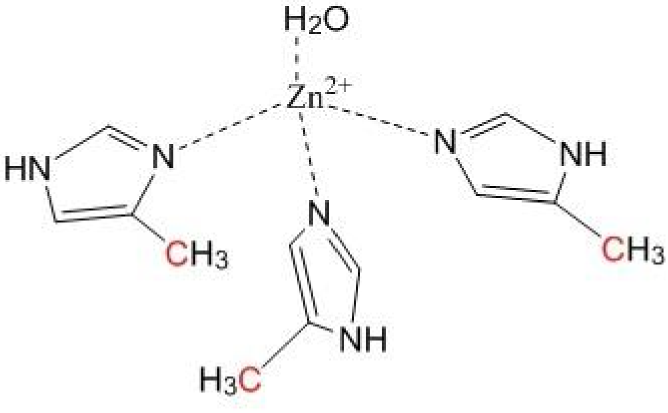
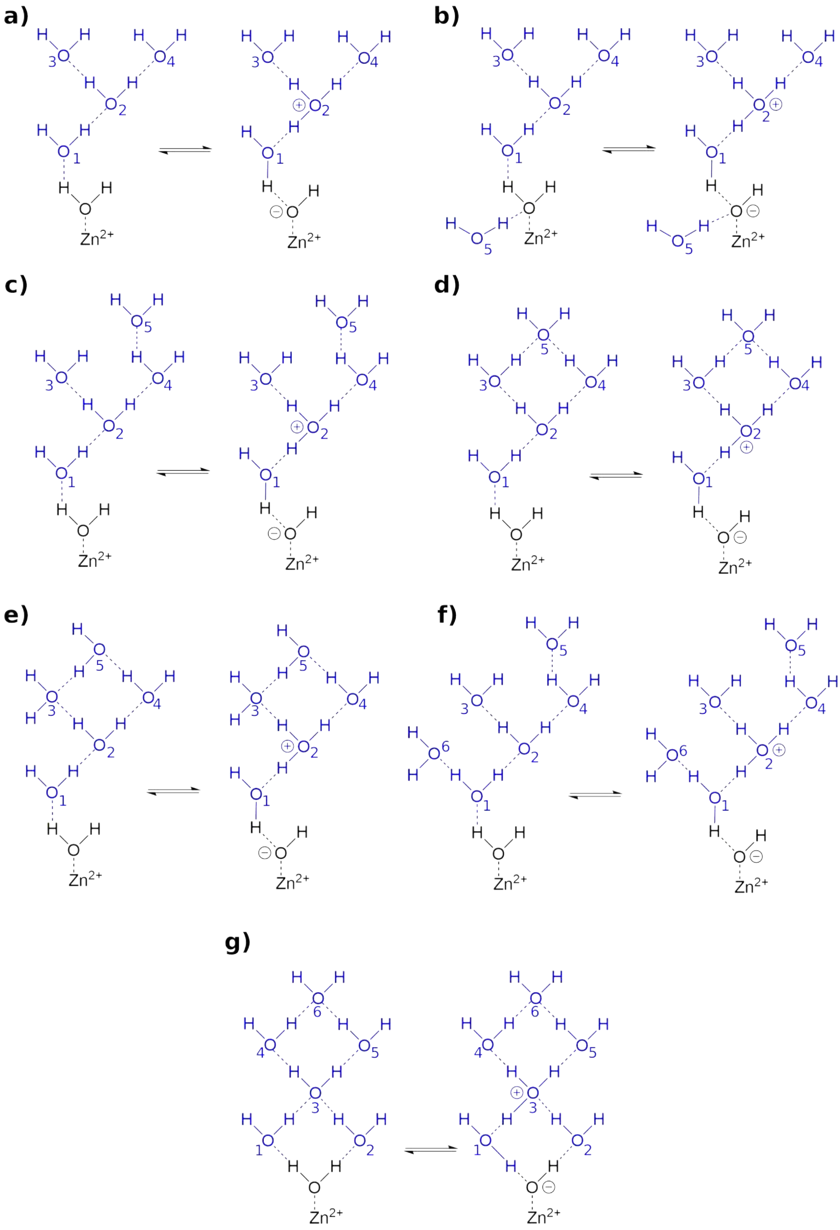


Figure 15.1: The model of the catalytic site, fixed carbons are marked in red

Initially, it has been supposed that a large network of water molecules could be per se sufficient to simulate both the proton carrier function of water and also its proton final acceptor role. Indeed different water “networks” with increasing number and disposition of water molecules (Figure 15.2 from **a** to **h**) have been tested but none of these allowed to model correctly the proton removal from coordinated water, actually the first (activation) step of catalysis. In fact, all attempted simulations always collapsed onto a single species, either the Zn-bound H_2O (starting materials) or the Zn-bound OH^- (products) in a mutually exclusive way.



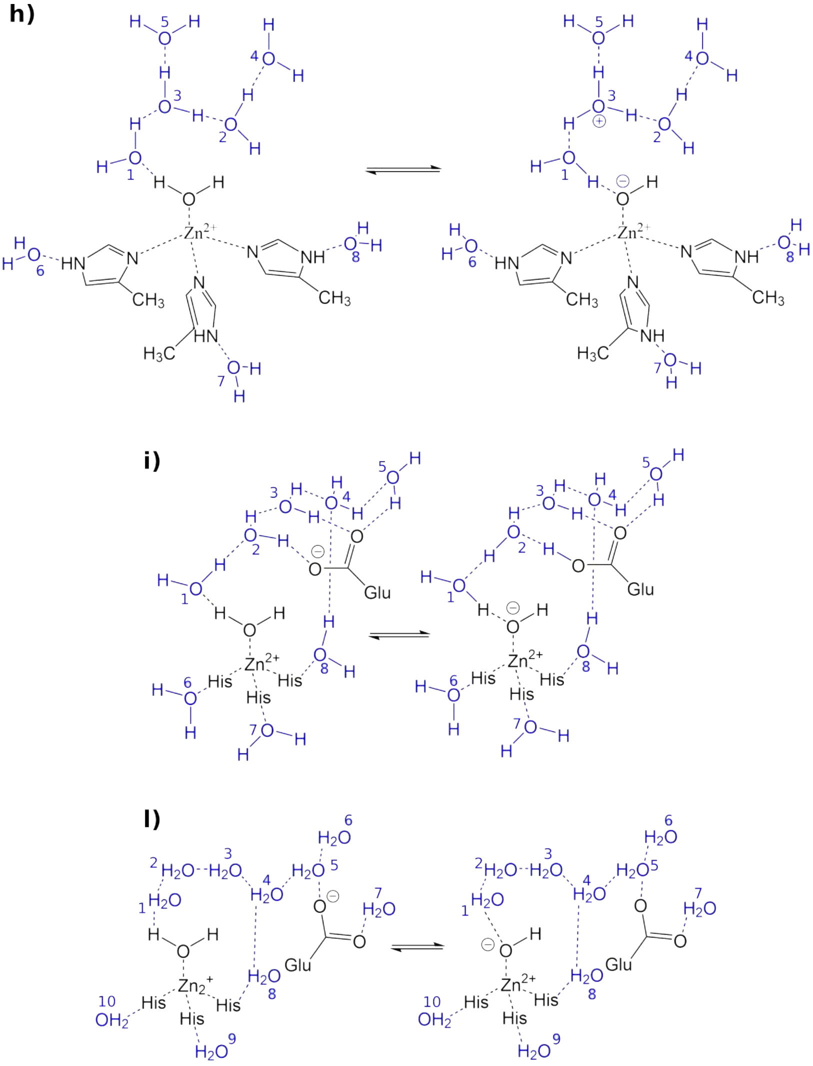


Figure 15.2: The panels show each water network tested. Simulation **(a)** has four water molecules but this configuration does not allow the formation of activated specie Zn-bound OH^-. In case **(b-e)** additional water molecules are present but the result is unchanged. In **(f)** and **(g)** case six water molecules are put in the model; in case of linear network we observed only the Zn-bound OH^- species, whereas by disposing a cyclic network of water molecules, only the Zn-bound H_2O is found. In **(h)** three water molecules are added close to histidines, one for each residue, (the hydrogen bond to the second nitrogen of the imidazole can tune proton transfer[135]). In simulations **(i)** and **(l)** a glutamate has been explicitly used as proton acceptor. In such two cases only the Zn-bound OH^- species can be obtained.

For this reason the computational model has been extended including a possible proton acceptor (Figure 15.2 **i** and **l**). In natural carbonic anhydrase a histidine or a aspartate work as proton acceptor [161], so we have looked for one of these residues in TRIL9CL23H. In particular a glutamate has been found at position 20^{th} on each chain at the same distance from Zn-bound water as in the carbonic β-anhydrase. Only the closest glutamate side chain (8.106 Å) has been included in the model. The working model consists of the zinc ion, the side chains of His_{23} and Glu_{20} and a water network made with 12 molecules, three of which are directly involved in proton transfer whereas the others actually play a role of solvent (Figure 15.3).

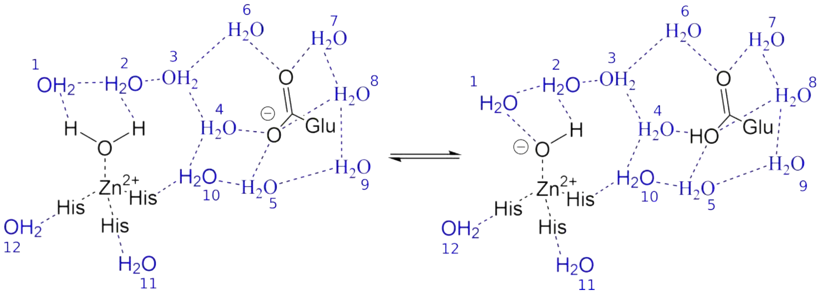


Figure 15.3: Model used for simulations. C_α of glutamate and C_β of histidine are fixed during each optimizations. Labels 2, 3 and 6 are for the bridging waters, i.e., they actually act as proton carrier, whereas the others are used for solvation waters. The net charge of the system is +1 (zinc=+2; his=0; glu=-1).

Chapter 16

Results and discussion

16.1 Characterization of catalytic pathway

16.1.1 B3LYP

The reaction pathway investigated for p-NPA ester hydrolysis by TRIL9CL23H is inspired to that well-established for carbonic anhydrase [152]. First we have studied the activation step, namely deprotonation of Zn-H_2O to Zn-OH^-. Calculations show a fully concerted proton transfer from Zn-H_2O to glutamate with the expected imaginary frequency associated to the transition normal mode of $544icm^{-1}$ (Figure 16.1 panel **b**). This step is exergonic and the free energy barrier is 3.7 kcal/mol, a value very close to that estimated for natural CA (2.2 kcal/mol) by kinetic measurement [162].

The p-NPA hydrolysis occurs according to a two-step mechanism through a tetrahedral intermediate formation on a C=O of ester which lies in between the two transition states TS-IV and TS-VI (Figure 16.2). In TS-IV (Figure 16.2 panel **b**) the formation of the $O_{hydroxyl}$-$C_{carbonyl}$ bond (1.709 Å) and of the $O_{carbonyl}$-Zn bond (2.146 Å) occur, while Zn-$O_{hydroxyl}$ distance increases to 2.094 Å (imaginary frequency of $198icm^{-1}$). The free energy barrier associated with this step is 9.7 kcal/mol. Results show that the intermediate V (Figure 16.2 panel **c**) is coordinated to zinc by carbonyl oxygen in a tetrahedral arrangement, with the $O_{carbonyl}$-Zn distance at 1.940 Å. TS-VI (Figure 16.2 panel **d**) is associated with the release of alcoholic portion, in particular results show a concerted mechanism which includes the break of $C_{carbonyl}$-$O_{alkoxide}$ bond (2.298 Å) and the H^+ transfer from Zn-bound acid to alkoxide (imaginary frequency: $87icm-1$). The free energy barrier is 6.7 kcal/mol and the concerted mechanism is consistent with the increase of acidity due to Zn, concomitantly with the strong base role played by alkoxide.

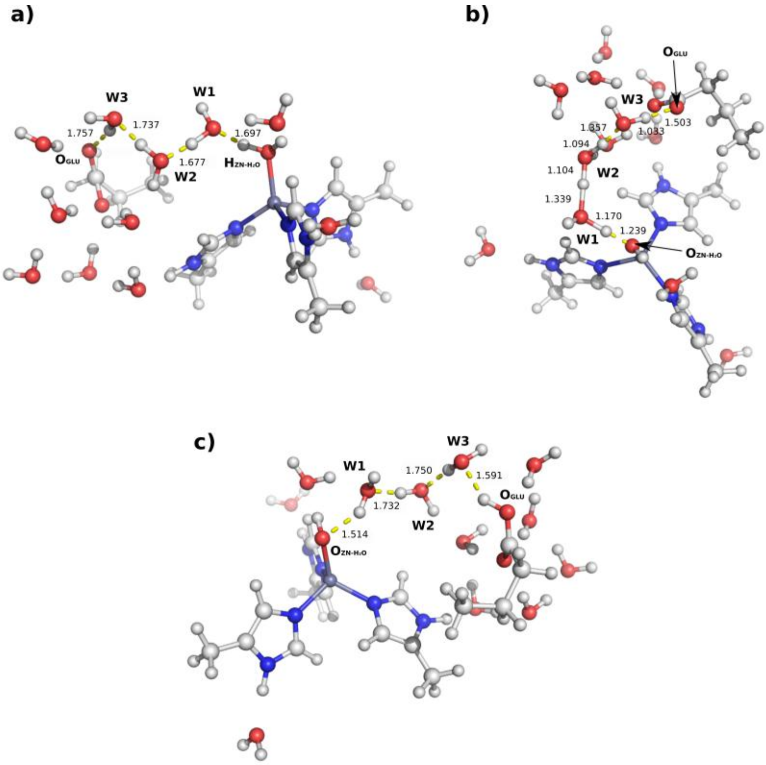


Figure 16.1: Activation of Zn-bound water. Color legend: oxygen in red, nitrogen in blue, zinc in violet, carbon in light gray, hydrogen in white.

Also the regeneration of active site occurs through a two-step mechanism and it leads to a five-coordinate intermediate like in the CA [152] (Figure 16.3). In TS-VIII (Figure 16.3 panel **b**) a water molecule comes close to the zinc ion and causes the formation of a distorted trigonal bipyramid geometry where water is in axial position (distance O$_{H_2O}$-Zn is 2.441 Å) whereas the acid is in equatorial position (distance O$_{acid}$-Zn is 1.980 Å). The free energy barrier is low on going from VII (Figure 16.3 panel **a**) to TS-VIII (Figure 16.3 panel **b**): 3.5 kcal/mol, (imaginary frequency is $17icm^{-1}$). The last step is the release/protonation of acid. During TSX (Figure 16.3 panel **d**) the Zn-O$_{acid}$ bond increases to 2.653 Å and a H^+ moves from glutamate to acid through water chain (imaginary frequency of $106icm^{-1}$). This step turned out to be the rate determining step of the entire catalytic process with an free energy barrier of 23.4 kcal/mol (Figure 16.4), in contrast with natural carbonic anhydrase where the rate limiting step is the proton transfer during water activation [136, 159, 128, 157, 151]. A possible reason is the different micro-environment in which the active sites of CA and TRIL9CL23H are embedded: in the latter there is no residue able to facilitate acid release. In XI the acid has been displaced by water molecule and active site is ready for a new catalytic cycle.

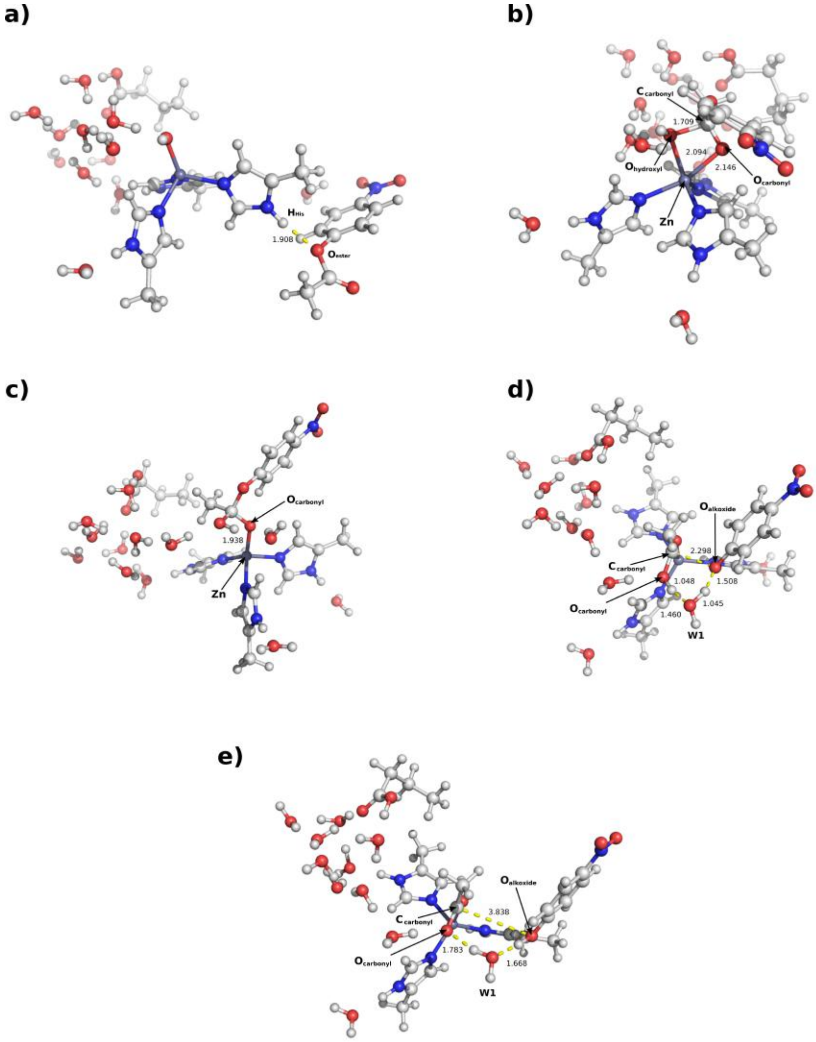


Figure 16.2: Ester hydrolysis and alcohol elimination. Color legend: oxygen in red, nitrogen in blue, zinc in violet, carbon in light gray, hydrogen in white.

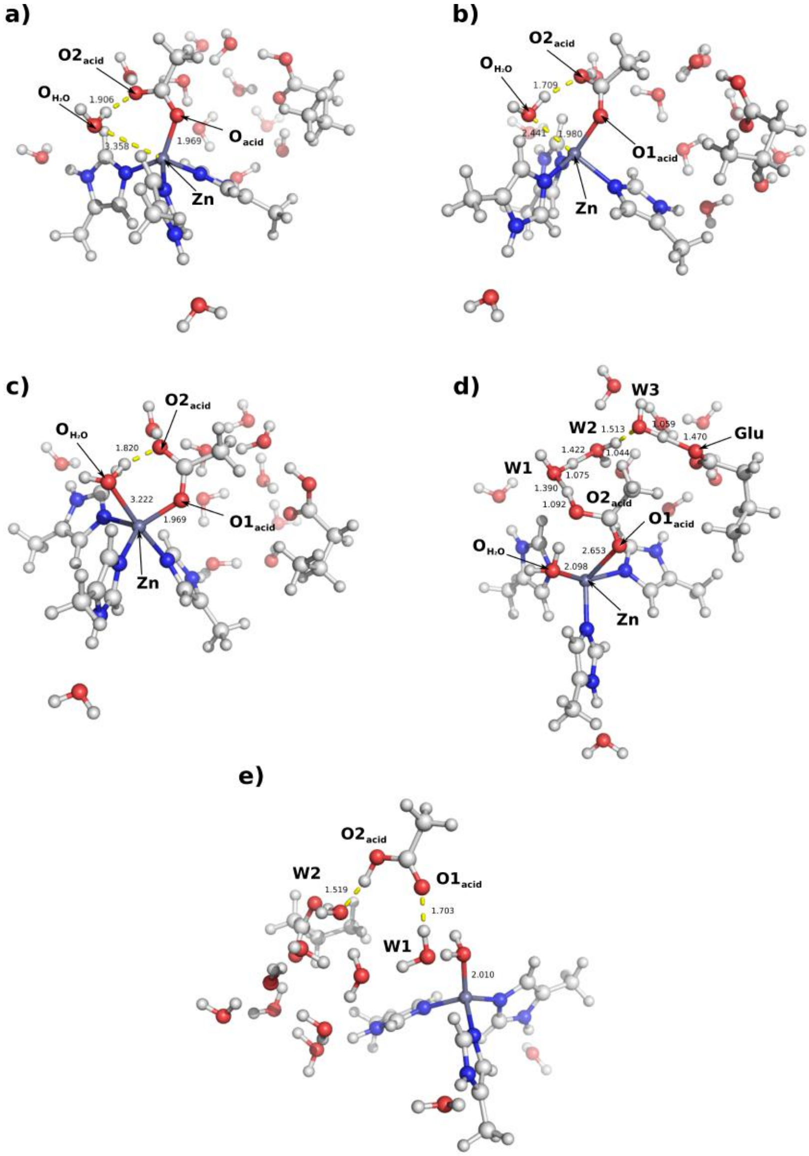
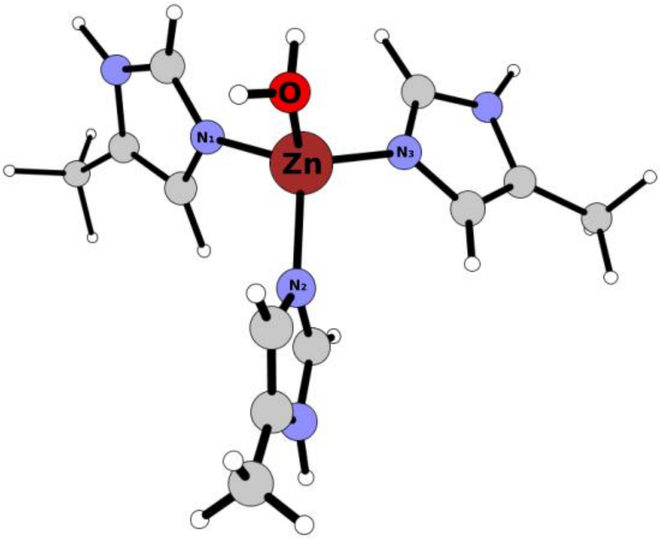


Figure 16.3: Acid release step and active site regeneration. Color legend: oxygen in red, nitrogen in blue, zinc in violet, carbon in light gray, hydrogen in white.

16.1.2 BP86

The entire catalytic pathway was also characterized using the pure correlation-exchange functional BP86 and the results were compared with simulations carried out with the hybrid functional B3LYP. Both functionals well reproduce the geometry of the complex as pointed out in table 16.1 in which are reported the main bonds and angles of Zn-(His)$_3$-H_2O active site. There are no significant differences in the distances monitored along the simulations (Table 16.2) and the two energetic landscapes have the same trend. In particular, also using the pure functional BP86, the rate-limiting step (12.2 kcal/mol) of the reaction is again the release of acid from the catalytic site (Figure 16.4).



		BP86	**B3LYP**	**VARIANCE**
BONDS (Å)	N_1-Zn	2.050	2.051	-0.001
	N_2-Zn	2.046	2.045	0.001
	N_3-Zn	2.017	2.020	-0.003
	Zn-O_{H_2O}	1.927	1.919	0.008
ANGLES (°)	N1-Zn-O_{H_2O}	109.6	107.6	1.935
	N_2-ZN-O_{H_2O}	119.3	117.0	2.224
	N_1-Zn-N_2	106.9	107.0	-0.880
	N_2-Zn-N_3	104.1	103.4	0.756

Table 16.1: Main bonds and angles of the Zn-(His)$_3$-H_2O evaluated using B3LYP and BP86 functionals.

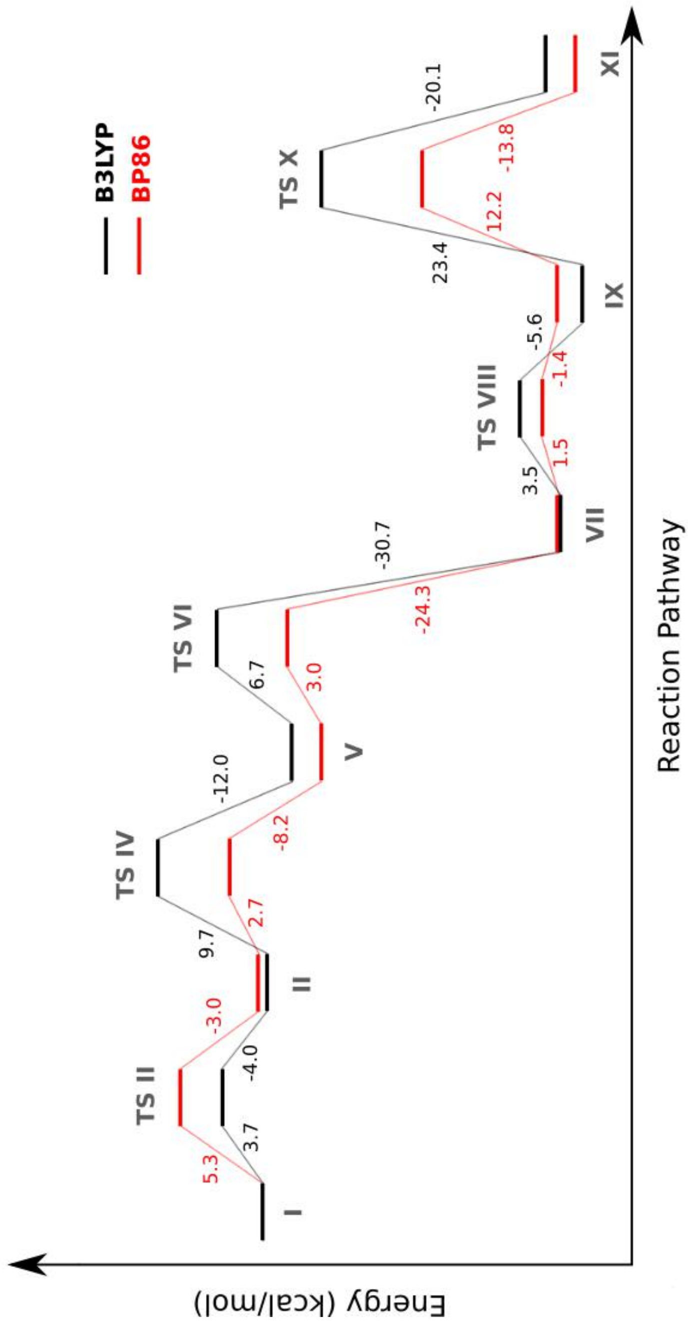


Figure 16.4: Energy profile obtained for the TRIL9CL23H catalytic process using B3LYP (black) and BP86 (red). Free energy values are expressed in kcal/mol.

		Å		
		B3LYP	**BP86**	**VARIANCE**
I	H_{Zn-H_2O}-O_{W1}	1.697	1.642	0.055
	H_{W1}-O_{W2}	1.677	1.587	0.090
	H_{W2}-O_{W3}	1.737	1.667	0.070
	H_{W3}-O_{Glu}	1.757	1.721	0.036
TS II	O_{Zn-H_2O}-H_{W1}	1.239	1.243	-0.004
	O_{Zn-H_2O}-H_{Zn-H_2O}	1.170	1.190	-0.020
	O_{W1}-H_{W2}	1.339	1.336	0.003
	H_{W2}-O_{W2}	1.104	1.125	-0.021
	O_{W2}-H_{W3}	1.094	1.150	-0.056
	H_{W3}-O_{W3}	1.357	1.300	0.057
	O_{W3}-H_{Glu}	1.033	1.072	-0.039
	H_{Glu}-O_{Glu}	1.503	1.437	0.066
III-a	O_{Zn-H_2O}-H_{W1}	1.514	1.483	0.031
	O_{W1}-H_{W2}	1.732	1.743	-0.011
	O_{W2}-H_{W3}	1.750	1.725	0.025
	O_{W3}-H_{Glu}	1.591	1.519	0.072
III-b	O_{ester}-H_{His}	1.908	2.043	-0135
TS IV	$O_{hydroxyl}$-$C_{carbonyl}$	1.709	1.758	-0.049
	$O_{carbonyl}$-Zn	2.146	2.228	-0.082
	Zn-$O_{hydroxyl}$	2.094	2.083	-0.011
V	$O_{carbonyl}$-Zn	1.940	1.938	0.002
TS VI	$C_{carbonyl}$-$O_{alkoxide}$	2.298	2.104	0.194
	$O_{carbonyl}$-H_a_{W1}	1.048	1.049	-0.001
	H_a_{W1}-O_{W1}	1.460	1.505	-0.045
	O_{W1}-H_b_{W1}	1.045	1.039	0.006
	H_b_{W1}-$O_{alkoxyde}$	1.508	1.549	-0.041
VII-a	$O_{carbonyl}$-H_a_{W1}	1.783	1.784	-0.001
	O_{W1}-H_{alcool}	1.668	1.695	-0.027
	$C_{carbonyl}$-$O_{alkoxide}$	3.838	3.896	-0.058

VII-b	Zn-O$_{acid}$	1.969	1.969	0
	O2$_{acid}$-O$_{H_2O}$	1.906	1.905	0.001
	O$_{H_2O}$-Zn	3.358	3.360	-0.002
TS VIII	O$_{H_2O}$-Zn	2.441	2.782	-0.341
	O$_{acid}$-Zn	1.980	1.992	-0.012
	O2$_{acid}$-O$_{H_2O}$	1.709	1.759	-0.050
IX	Zn-O1$_{acid}$	1.969	2.003	-0.034
	O2$_{acid}$-O$_{H_2O}$	1.820	1.687	0.133
	Zn-O$_{H_2O}$	3.222	2.385	0.837
TS X	Zn-O$_{acid}$	2.653	2.653	0
	Zn-O$_{H_2O}$	2.098	2.138	-0.040
	O2$_{acid}$-H$_{W1}$	1.092	1.090	0.002
	H$_{W1}$-O$_{W1}$	1.390	1.395	-0.005
	O$_{W1}$-H$_{W2}$	1.075	1.083	-0.008
	H$_{W2}$-O$_{W2}$	1.422	1.413	0.009
	O$_{W2}$-H$_{W3}$	1.044	1.113	-0.069
	H$_{W3}$-O$_{W3}$	1.513	1.538	-0.025
	O$_{W3}$-H$_{Glu}$	1.059	1.174	-0115
	H$_{Glu}$-O$_{Glu}$	1.470	1.399	0.071
XI	Zn-O$_{H_20}$	2.010	1.973	0.037
	O1$_{acid}$-H$_{W1}$	1.703	1.644	0.059
	O2$_{acid}$-O$_{W2}$	1.519	1.554	-0.035

Table 16.2: Main distances monitored along the simulations

16.2 Effect of Leu$_{19}$→Thr

A larger sized model that includes the second coordination sphere around the active site has been subsequently used to study the effect of substitution of $Leu19$ to Thr. In α-CA, Thr_{199} is essential for activity[152] and occupies the same position of Leu_{19}. Zn-$H_2O(Leu)_3$ and Zn-$H_2O(Thr)_3$ have been thus used as computational models to evaluate the effects of the mentioned mutation (Figure 16.5). In both systems the activation energy of proton transfer and the pK_a have been calculated for each Zn-bound water proton. All the possible pathways conceivable for proton transfer have been investigated. Unexpectedly, the Leu/Thr substitution has a negative effect on the catalysis, as the barrier energy of proton transfer and computed pK_a in Zn-$H_2O(Thr)_3$ are significantly higher compared to the Zn-$H_2O(Leu)_3$ (Table 16.3). In natural CA, Thr_{199} is essential, along with Glu_{106}, to orient the nucleophilic zinc-bound hydroxide for optimal attack on the substrate.[150, 164] The interaction between threonine and glutamate is fairly well preserved in CA[160], in particular Glu_{106} acts as hydrogen-bond acceptor in its interaction with Thr_{199} (Figure 16.6). Thus, the absence of glutamate in the de-novo designed system herein investigated, is fully compatible with a relative detriment of acidity by protons of the Zn-H_2O moiety.[148] Moreover, our results show that threonine cannot be involved in the proton abstraction and generation of nucleophilic Zn-OH^-, as the energy barrier associated to this step is much higher than in the wild-type system.

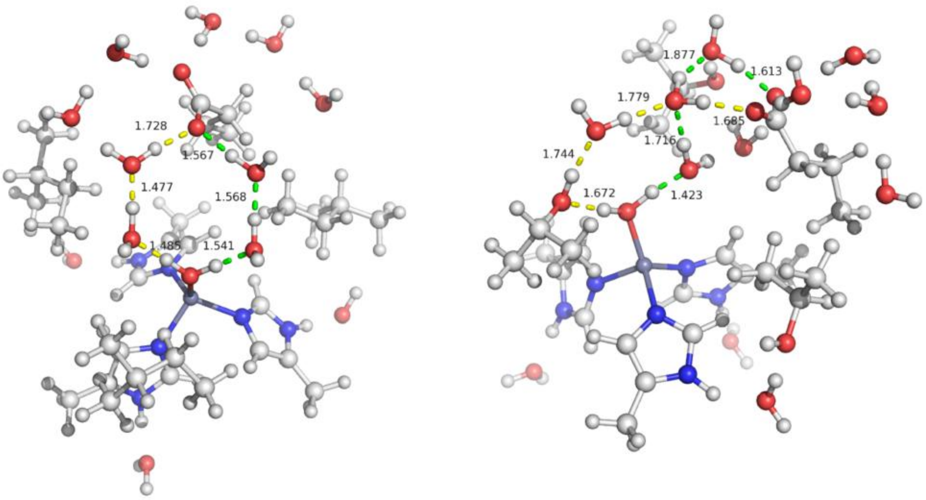


Figure 16.5: Structure of the model Zn-$H_2O(Leu)_3$ (on the left) and model $H_2O(Thr)_3$ (on the right). Green and yellow dashed lines pinpoint the proton transfer route.

MODEL	$\Delta E_{reagent-product}$	$\Delta E^{\ddagger}_{forward}$	$\Delta E^{\ddagger}_{backward}$	pK_a
Zn-$H_2O(Leu)_3$ [proton a]	0.2	0.6	-0.4	4.7
Zn-$H_2O(Leu)_3$ [proton b]	2.5	2.7	-0.2	4.5
Zn-$H_2O(Thr)_3$ [proton a]	8.3	9.9	-1.6	8.3
Zn-$H_2O(Thr)_3$ [proton b]	11.2	14.4	-3.2	11.2

Table 16.3: Summary of $\Delta E_{reagent-product}$,$\Delta E^{\ddagger}_{forward}$,$\Delta E^{\ddagger}_{backward}$ and pK_a values evaluated for Zn-$H_2O(Leu)_3$ and Zn-$H_2O(Thr)$ during water activation step. Energy values are reported in kcal/mol.

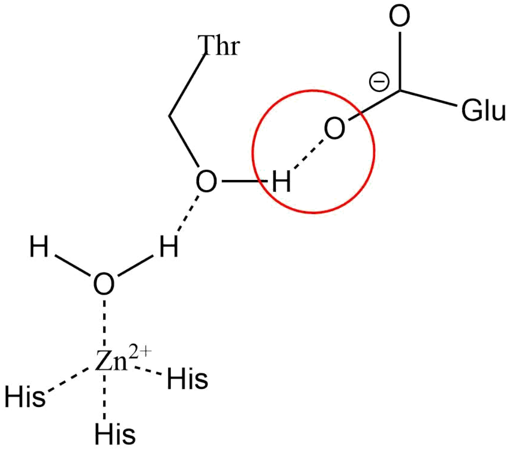
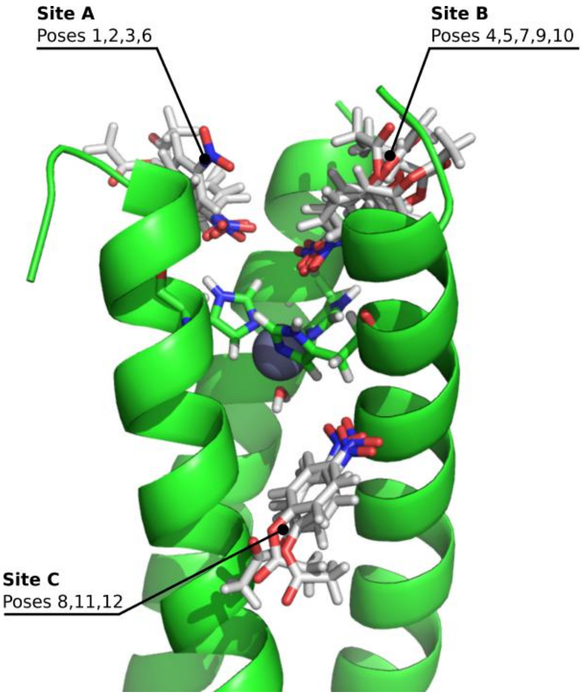


Figure 16.6: The hydrogen bond network among Zn-H_2O, Thr_{199} and Glu_{106} in natural carbonic anhydrase. In TRL9CL23H there is no glutamate residue and threonine is not able to assume a “correct” orientation (i.e., CA-like).

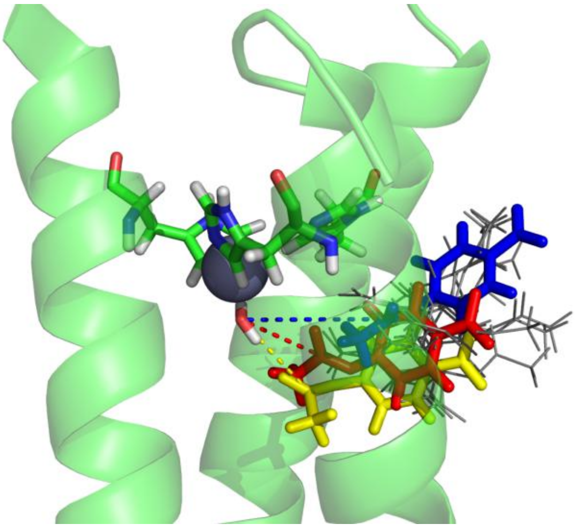
16.3 Docking p–NPA to TRIL9CL23H

Molecular Docking simulations (see Methods for details) have been carried out aimed to find one (or more) possible binding site(s) for p-nitrophenyl acetate on the protein side. First, the structure of TRIL9CL23H from 3PBJ pdb file has been partially optimized (only the side chains have been subject to energy minimization). Then, the optimized structure has been used as target for a rigid docking simulation whit p-NPA as ligand: results show three putative binding sites (Figure 16.7). Two of these (Sites A and B) are slightly energetically preferred over all others and are located in regions that are at the opposite side respect to the water bound to Zn in the active site. Such unfavorable orientation of NPA (vs Zn-H_2O) would suggest that binding to sites A and B could play a detrimental role for catalysis. Otherwise stated, a fraction of substrate could be “wasted” in binding to such two regions. This result could represent another point underlying the less catalytic efficiency of TRIL9CL23H vs CA (that evidently must have evolved so as to avoid aforementioned wasting-spots). However, three other ligand conformations have been identified on the side envelope of TRIL9CL23H (Site C) that are very close (about 1 kcal/mol) in energy to those in Sites A and B. Remarkably, they could be more directly involved in catalysis as NPA in this case is located from the same side of Zn-H_2O. Among conformations found for Site C, it has been selected that featuring the best score (Figure 16.7 ligand pose n. 8) as starting point for IDF simulations. Although site C has positional features that are more compatible with catalysis events, it is still an exterior region of the peptide envelope. Therefore IDF simulations have been initially started out by keeping the interaction between the ester carbonyl and the Zn-bound water restrained in order to verify the possibility for the ligand to move to the interior of the protein envelope. The simulation has showed three interesting “inner” ligand binding modes (Figure 16.8 structures in red, yellow and blue), that proved energetically favored and also feature an orientation which is coherent with nucleophilic attack by the Zn-OH^- moiety. Among side poses that proximal to active site, the closest one (Figure 16.8 red structures) has selected for an optimization refinement. Two energy minimizations have been carried out (in all simulations backbone was fixed to prevent unfolding); during the first one, NPA has been kept fixed and only side chains have been relaxed. The second simulation featured both NPA and protein side chains free to rearrange (Figure 16.9). Results show that NPA is located 4.779 Å from Zn-bound water and it makes a contact with Lys_{22} (chain D) and Glu_{20} (chain E) through NO_2. The found configuration is representative of a state that may occur before the nucleophilic attack, yet it is relevant because it demonstrates the energetic accessibility of pre-catalytic interactions between protein and substrate.



Ligand Pose	(G)Score kcal/mol
1	-4.3
2	-4.0
3	-3.9
4	-3.9
5	-3.4
6	-3.3
7	-3.2
8	-3.1
9	-3.1
10	-2.9
11	-2.8
12	-2.7

Figure 16.7: Results of rigid docking simulation. In Table are reported the docking score in ascending order for each pose. Marked are the three poses closer to the active site. Pose 8 is the ligand pose selected for IDF.



Pose	(G)Score kcal/mol	Distance (Å) O_{H_2O}- C=$O_{p\text{-}NPA}$
1 (yellow)	-5.5	4.245
2	-4.8	//
3	-4.7	//
4 (red)	-4.7	3.924
5	-4.5	//
6 (blue)	-4.4	6.122
7	-4.3	//
8	-4.2	//
9	-4.1	//
10	-4.0	//
11	-3.7	//
12	-3.6	//

Figure 16.8: Result of IDF simulation. In table are reported docking scores (decreasing order of stability) for each pose; marked are those poses that are closer to the Zn-bound water.

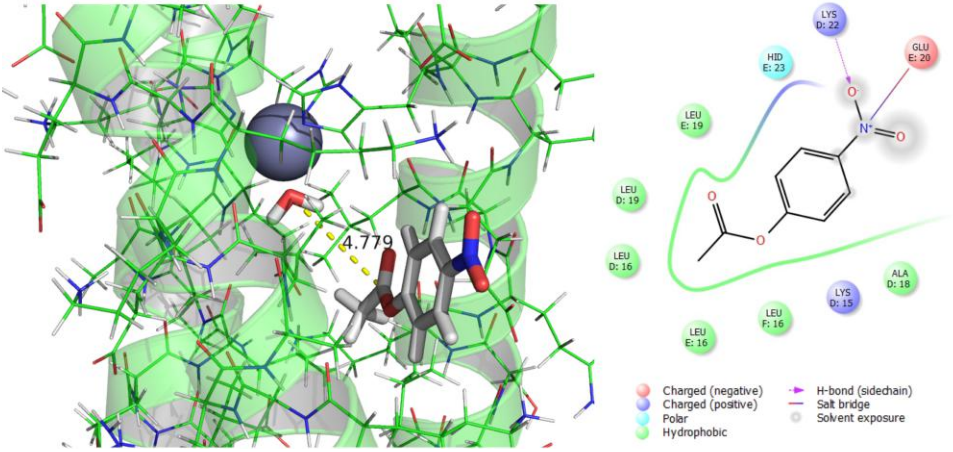


Figure 16.9: Result of structure optimization and a docking model structure of p-NPA into the 3PBJ binding pocket .

Chapter 17

Conclusion

In conclusion DFT has been used to investigate viable pathway of NPA hydrolysis as catalyzed by TRIL9CL23H, a *de-novo* designed peptide, inspired to carbonic anhydrase. According to results, the catalytic mechanism is qualitatively similar to the natural carbonic anhydrase, with the interesting difference consisting in the rate-determining step. Indeed, in contrast to CA where the rate-determining step is water activation leading from Zn-H_2O to Zn-OH^-, in the case of TRIL9CL23H the regeneration of the active site (the acid release), features the highest activation energy. A conceivable rationale for such outcome is the absence of any residue able to assist the acid release from active site. Another important observation is the detrimental effect that mutation $Lue_{19} \rightarrow Thr$ has on catalysis. Partially unexpectedly, even if a Thr is essential in the natural enzyme, in the case of *de-novo* designed system, this residue leads to an increase of the pK_a of Zn-H_2O; this fact underlines the importance of the protein environment and how it is able to modify and tune also established properties of a given residue. Actually, the impossibility to predict intermolecular interactions is a limitation in *de novo* design of catalytic proteins [143, 120]. Computational studies are helpful in such sense, as they can provide rationalization of unexpected results. Specifically in the present study, the absence of a Glu in TRIL9CL23H has been recognized as the crucial factor justifying the detrimental effect of the Leu-Thr mutation. A Glu is in fact a Thr partner in the natural CA in relevant (and conserved) orienting/hydrogen bonding functions. Finally, we have searched for possible binding sites for NPA to TRIL9CL23H . It has been identified one inner binding site (Figure 16.8) close to the Zn-bound water. Docking simulations predict that NPA has binding affinity also for sites that are located on the surface side of the peptide: two (Figure 16.7 sites A and B) are at the C-terminal end and appear incompatible for catalysis (thus representing another possible reason justifying a slightly less efficient process vs natural CA). Another outer region (Figure 16.7 site C) is instead located on the same side of the Zn bound water, thus possibly representing a region where ligand is easily accommodated before entering catalytic site.

Bibliography

[119] Reinhart Ahlrichs, Michael Bär, Marco Häser, Hans Horn, and Christoph Kölmel. Electronic structure calculations on workstation computers: The program system turbomole. *Chemical Physics Letters*, 162(3):165–169, 1989.

[120] Lars Baltzer and William F DeGrado. Engineering and design. *Current Opinion in Structural Biology*, 14(4):455–457, aug 2004.

[121] J A Y L Banks, Hege S Beard, Yixiang Cao, A R T E Cho, Wolfgang Damm, Ramy Farid, Anthony K Felts, Thomas A Halgren, Daniel T Mainz, J O N R, Robert Murphy, Dean M Philipp, Matthew P Repasky, Linda Y, Bruce J Berne, Richard A Friesner, Emilio Gallicchio, and M Levy. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *Journal of computational chemistry*, 26(16):1752–1780, 2005.

[122] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, 38(6):3098–3100, 1988.

[123] Axel D. Becke. Density-functional thermochemistry.III. The role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648, 1993.

[124] F. Bernardi, Andrea Bottoni, Marco De Vivo, Marco Garavelli, Gyorgy Keseru, and Gabor Naray-Szabo. A hypothetical mechanism for HIV-1 integrase catalytic action: DFT modelling of a bio-mimetic environment. *Chemical Physics Letters*, 362(1-2):1–7, aug 2002.

[125] Art D. Bochevarov, Edward Harder, Thomas F. Hughes, Jeremy R. Greenwood, Dale A. Braden, Dean M. Philipp, David Rinaldo, Mathew D. Halls, Jing Zhang,

and Richard A. Friesner. Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *International Journal of Quantum Chemistry*, 113(18):2110–2142, 2013.

[126] D N Bolon and S L Mayo. Enzyme like proteins by computational design. *Proceedings of the National Academy of Sciences of the United States of America*, 98(25):14274–9, dec 2001.

[127] Christopher J Cramer. *Essentials of Computational Chemistry*. Wiley, second edi edition, 2004.

[128] Qiang Cui and Martin Karplus. Is a Proton Wire Concerted or Stepwise? A Model Study of Proton Transfer in Carbonic Anhydrase. *The Journal of Physical Chemistry B*, 107(4):1071–1078, jan 2003.

[129] W F DeGrado, C M Summa, V Pavone, F Nastri, and a Lombardi. De novo design and structural characterization of proteins and metalloproteins. *Annual review of biochemistry*, 68:779–819, jan 1999.

[130] J R Desjarlais and N D Clarke. Computer search algorithms in protein modification and design. *Current opinion in structural biology*, 8(4):471–5, aug 1998.

[131] Patricia C Dos Santos, Suzanne M Mayer, Brett M Barney, Lance C Seefeldt, and Dennis R Dean. Alkyne substrate interaction within the nitrogenase MoFe protein. *Journal of inorganic biochemistry*, 101(11-12):1642–8, nov 2007.

[132] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2):1007–1023, 1989.

[133] Christopher B Eiben, Justin B Siegel, Jacob B Bale, Seth Cooper, Firas Khatib, Betty W Shen, Foldit Players, Barry L Stoddard, Zoran Popovic, and David Baker. Increased Diels Alderase Activity through Foldit Player Guided Backbone Remodeling. *Nature Biotechnology*, 30(2):190–192, 2012.

[134] Richard a Friesner, Jay L Banks, Robert B Murphy, Thomas a Halgren, Jasna J Klicic, Daniel T Mainz, Matthew P Repasky, Eric H Knoll, Mee Shelley, Jason K Perry, David E Shaw, Perry Francis, and Peter S Shenkin. Glide: a new approach for

rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry*, 47(7):1739–49, mar 2004.

[135] Gilles Frison and Gilles Ohanessian. Metal-histidine-glutamate as a regulator of enzymatic cycles: a case study of carbonic anhydrase. *Physical chemistry chemical physics : PCCP*, 11(2):374–83, 2009.

[136] Ameen F Ghannam, William Tsen, and Roger S Rowletts. THE JOURNAL OF BIOLOGICAL CHEMISTRY Activation Parameters for the Carbonic Anhydrase 11-catalyzed Hydration of C02. 261(3):1164–1169, 1986.

[137] Stefan Grimme. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry*, 25(12):1463–1473, sep 2004.

[138] Stefan Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry*, 27(15):1787–1799, nov 2006.

[139] Thomas a Halgren, Robert B Murphy, Richard a Friesner, Hege S Beard, Leah L Frye, W Thomas Pollard, and Jay L Banks. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of medicinal chemistry*, 47(7):1750–9, mar 2004.

[140] Matthew P. Jacobson, Richard A. Friesner, Zhexin Xiang, and Barry Honig. On the role of the crystal environment in determining protein side-chain conformations. *Journal of Molecular Biology*, 320(3):597–608, 2002.

[141] Matthew P. Jacobson, David L. Pincus, Chaya S. Rapp, Tyler J.F. Day, Barry Honig, David E. Shaw, and Richard A. Friesner. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins: Structure, Function and Genetics*, 55(2):351–367, 2004.

[142] Frank Jensen. *Introduction to Computational Chemistry*. Wiley, second edi edition, 2007.

[143] J Kaplan and W F DeGrado. De novo design of catalytic proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 101(32):11566–70, aug 2004.

[144] Rick A. Kendall, Thom H. Dunning, and Robert J. Harrison. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of Chemical Physics*, 96(9):6796–6806, 1992.

[145] K S Kim, P Tarakeshwar, and J Y Lee. Molecular Clusters of pi-Systems: Theoretical Studies of Structures, Spectra, and Origin of Interaction Energies. *Chemical reviews*, 100(11):4145–86, nov 2000.

[146] A Klamt and G Schuurmann. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. (2), 1993.

[147] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2):785–789, 1988.

[148] Zhiwei LIANG, Yafeng XUE, Gity BEHRAVAN, Bengt-Harald H JONSSON, and Sven LINDSKOG. Importance of the conserved active-site residues Try7, Glu106 and Thr199 for the catalytic function of human carbonic anhydrase II. *European Journal of Biochemistry*, 211(3):821–827, 1993.

[149] James T MacDonald, Katarzyna Maksimiak, Michael I Sadowski, and William R Taylor. De novo backbone scaffolds for protein design. *Proteins*, 78(5):1311–25, apr 2010.

[150] K. M. Merz Jr. Insights into the function of the zinc hydroxide-Thr199-Glu106 hydrogen bonding network in carbonic anhydrases. *Journal of Molecular Biology*, 214(4):799–802, aug 1990.

[151] Rose Mikulski, Balendu Sankara Avvaru, Chingkuang Tu, Nicolette Case, Robert McKenna, and David N Silverman. Kinetic and crystallographic studies of the role of tyrosine 7 in the active site of human carbonic anhydrase II. *Archives of biochemistry and biophysics*, 506(2):181–7, feb 2011.

[152] Gian Pietro Miscione, Marco Stenta, Domenico Spinelli, Ernst Anders, and Andrea Bottoni. New computational evidence for the catalytic mechanism of carbonic anhydrase. *Theoretical Chemistry Accounts*, 118(1):193–201, apr 2007.

[153] Robert J Pantazes, Matthew J Grisewood, and Costas D Maranas. Recent advances in computational protein design. *Current opinion in structural biology*, 21(4):467–72, aug 2011.

[154] John P Perdew. f (3) ’. (June):8822–8824, 1986.

[155] G. A. Petersson and Mohammad A. Al-Laham. A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. *The Journal of Chemical Physics*, 94(9):6081–6090, 1991.

[156] G. A. Petersson, Andrew Bennett, Thomas G. Tensfeldt, Mohammad A. Al-Laham, William A. Shirley, and John Mantzaris. A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements. *The Journal of Chemical Physics*, 89(4):2193–2218, 1988.

[157] Demian Riccardi, Shuo Yang, and Qiang Cui. Proton transfer function of carbonic anhydrase: Insights from QM/MM simulations. *Biochimica et biophysica acta*, 1804(2):342–51, feb 2010.

[158] Ansgar Schäfer, Christian Huber, and Reinhart Ahlrichs. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *The Journal of Chemical Physics*, 100(8):5829–5835, 1994.

[159] Lindskog S. Silverman D. N. The catalytic mechanism of carbonic anhydrse: implications of a rate limiting proteolysis of water. *Accounts of Chemical Research*, 21:30–36, 1988.

[160] Mera K. M. Toba S., Colombo G. Toba 1999.pdf. *Journal of the Amaerican Chemical Society*, 121:2290–2302, 1999.

[161] B. C. Tripp. Carbonic Anhydrase: New Insights for an Ancient Enzyme. *Journal of Biological Chemistry*, 276(52):48615–48618, nov 2001.

[162] C Tu, M Qian, J N Earnhardt, P J Laipis, and D N Silverman. Properties of intramolecular proton transfer in carbonic anhydrase III. *Biophysical journal*, 74(6):3182–9, jun 1998.

[163] David E. Woon and Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *The Journal of Chemical Physics*, 98(2):1358–1371, 1993.

[164] Zhidong Xue, Dong Xu, Yan Wang, and Yang Zhang. ThreaDom: extracting protein domain boundary information from multiple threading alignments. *Bioinformatics (Oxford, England)*, 29(13):i247–i256, jul 2013.

[165] Vincent L. Zastrow, Melissa L., Peacock, Anna F. A., Stuckey, Jeanne A., Pecoraro. Hydrolitic catalysis and structural stabilization in a designed metalloprotein. *Nature Chemistry*, 4(2):118–123, 2012.

[166] Tom. Ziegler. Approximate density functional theory as a practical tool in molecular energetics and dynamics. *Chemical Reviews*, 91(5):651–667, jul 1991.

Part IV

Investigation of the the dynamic feature of different mutants of Mre11 subunit of MRX complex

Chapter 18

Abstract

Homologous recombination is triggered by nucleolytic degradation (resection) of DNA double-strand breaks (DSBs). DSB resection requires the Mre11-Rad50-Xrs2 (MRX) complex, which promotes the resection activity of Exo1 nuclease through a poorly understood mechanism. This work describes the Mre11-R10T mutant variant that accelerates DSB resection compared to wild type Mre11 by potentiating Exo1 resection activity. Molecular dynamics simulations show that rotation of Mre11 capping domains is actually able to induce dsDNA to unwind. The R10T substitution causes an altered rotation orientation of the Mre11 capping domain that leads to persistent melting of the duplex DNA. MRX creates a specific DNA structure that promotes Exo1 resection activity by facilitating the persistence of this nuclease on the DSB ends, uncovering a novel MRX function in DSB resection.

Chapter 19

Introduction

DNA double-strand breaks (DSBs) can be repaired by non-homologous end joining (NHEJ), which comprises direct religation of the DSB ends, or homologous recombination (HR), which uses intact homologous duplex DNA sequences (sister chromatids or homologous chromosomes) as a template for repair [178, 191]. Once a DSB occurs, the highly evolutionarily conserved MRX (Mre11-Rad50-Xrs2) complex and the Ku70/Ku80 (Ku) heterodimer are recruited to DNA ends independently to each other and play an important role in the choice between NHEJ and HR [190, 204]. Ku binding to DNA ends mediates recruitment of downstream NHEJ factors [189, 177, 206, 179]. MRX, together with Sae2, in turn promotes the nucleolytically processing of the 5' ends of the DSB in a process called resection [199, 201]. When resection takes place, the generation of 3'-ended ssDNA tails inhibits NHEJ and channels DSB repair into HR. In both yeast and mammals, the current model for resection is that Sae2 stimulates Mre11 endonuclease activity, which generates a nick in the 5'-terminated strands at both DSB ends. This endonucleolytic cleavage provides the access for Exo1 and Dna2 nucleases capable of degrading DNA in a 5'-3' direction [192, 207, 174, 197, 173]. While Exo1 degrades the 5'-terminated strands within duplex DNA [200], Dna2 cleaves ssDNA overhangs adjoining a duplex DNA and this processing activity requires the RecQ helicase Sgs1 (BLM in humans), which unwinds double-stranded DNA (dsDNA) in a 3'-5' direction [188, 207, 197, 174]. The endonucleolytic cleavage catalyzed by MRX-Sae2 is particularly important to initiate resection at DSBs whose DNA ends are not accessible to Exo1 and Sgs1-Dna2 due to the presence of chemical modifications or covalently bound-proteins [199, 201]. However, the lack of any MRX subunit, but not of Mre11 nuclease activity, reduces the resection efficiency of DSBs with free 3' hydroxyl and 5' phosphate groups [180], indicating that MRX is important even at these DSBs to potentiate Exo1 and/or Sgs1-Dna2 activity. Consistent with these observations, biochemical reconstitution experiments have shown that MRX/MRN enhances the ability of Sgs1/BLM to unwind DNA independently of Mre11 nuclease activity possibly by increasing Sgs1/BLM affinity for DNA ends [174, 195, 196, 197, 172]. Furthermore, MRX/MRN enhances both the affinity to DNA ends of Exo1 and its intrinsic processivity activity [174, 195, 196, 197, 172]. These results are consistent with *in vivo* data that indicate a requirement for MRX in the recruitment of both Sgs1 and Exo1 to DNA ends [198]. On the other hand, resection catalyzed by Exo1 is inhibited by the presence of the Ku complex, which forms a ring-shaped molecule that binds DNA ends [202]. The absence of Ku suppresses the resection defect of mre11Δ and sae2Δ cells in an Exo1-dependent manner [193, 183], suggesting that MRX promotes Exo1 action by removing Ku from DNA ends. As Ku preferentially binds double-stranded DNA ends over ssDNA [194, 184, 171, 182, 202], the MRX-Sae2 clipping could counteract Ku association to DNA ends by creating a substrate less suitable for Ku binding [193, 175]. Alternatively, since the absence of MRX, but not of Sae2 or Mre11 nuclease activity, causes an increase of Ku association at DNA ends [206, 204, 198], MRX could limit Ku binding by competing with it for binding the same DNA end. Here, we describe the mre11-R10T allele that suppresses the DNA damage sensitivity and the resection defect of sae2Δ cells. Mre11-R10T not only bypasses Sae2 function in DSB resection, but also accelerates the resection process compared to wild type Mre11. This enhanced DSB resection is due mainly to the activity of Exo1, whose association to DSBs is increased in mre11-R10T cells. As a consequence, the increased Exo1 activity leads to a decreased Ku association to DSBs and an enhanced DSB resection in G1, indicating that Exo1 can counteract Ku association to DSBs. Molecular dynamic simulations show that the DNA ends interact with the two capping domains of the Mre11 dimer and the rotation of these domains results in partial unwinding of the DNA double helix. The Mre11-R10T variant exhibits an altered positioning of the capping domains that leads to a persistent melting of the dsDNA end, suggesting that Mre11-R10T enhances Exo1-mediated DSB resection by providing a DNA structure that results in a higher affinity binding for Exo1.

Chapter 20

Computational methods

20.1 Homology modeling

A BLASTX search of the Protein Data Bank (PDB) using ScMre11 protein sequence as a query showed high sequence identity (52%) with SpMre11 (PDB ID: 4FBW) and also (50%) with CtMre11 (PDB ID: 4YKE) (see Appendix A). A homology based model of ScMre11 protein spanning endonuclease and capping domains (1-416 aa) was constructed based on the crystal structure coordinates of 4FBW. The pair wise alignment was performed using the homology module of PRIME [187, 186]. The pair wise alignment was improved manually by minor editing based on the secondary structure predictions as well, and the energy-driven modeling option was used in order to allow energy minimization using the force-field OPLS-2005[169]. The initial structure for Rad50 RBD, spanning amino acids 443-526, was obtained by homology modeling based on the crystal coordinates of the Rad50-binding domain of Mre11 in complex with Rad50 from *C. thermophilum* (PDB ID: 5DA9). Since the sequence identity between the two proteins is limited in this domain, the pair wise alignment obtained using the PRIME homology module was improved manually by minor editing based on the secondary structure predictions and imposing conservation of the relevant Rad50 interacting surface residues. The disordered connecting regions between the first model and the RBD model were designed ex novo by the energy-driven PRIME homology module and then refined by PRIME loops refinement module. The whole model of ScMre11, spanning amino acids 1-416, was then processed for molecular dynamics as a heterodimer with Rad50. A model for ScRad50 globular domain and initial coiled-coil regions, bound to ATP, was obtained by homology modeling as described above based on the crystal coordinates of CtRad50 (PDB ID: 5DA9), sharing a 58% identity with ScRad50. An initial model of the tetrameric structure of Mre11-Rad50 complex was obtained by superposition of ScMre11 and ScRad50 (ATP-bound) on the tetrameric complex structure from *M. jannaschii* (PDB ID: 5F3W). This structure was then refined by molecular dynamics (see Appendix B).

20.2 Molecular dynamics

MD simulations were performed using GROMACS 5.0.4 software package [185, 167] and the AMBER99 force field [203]. During the simulations, the interactions between magnesium ions (Mg^{2+}) and their coordinating residues were treated using non-bonded model with van der Waals and electrostatic terms present in the AMBER99 force field. Protein structures were soaked in a cubic box of SPC/E (Extended Single Point Charge) [170, 176] water molecules and simulated using periodic boundary conditions. All protein atoms are at the distance of 1 nm from the box edges. The ionization state of residues was set to be consistent with physiological pH (7.0 pH ± 0.2). Na^+ ions were added to all molecular systems as counterions to obtain electroneutral models (12 Na^+ were added to the wild type Mre11 system and 13 Na^+ to the mutant Mre11 system). All systems were relaxed trough a steepest descent minimization with a limit of 50000 cycles. Temperature and pressure were equilibrated at 300K and 1 atm with a NVT (isothermal-isochoric) simulation of 4 ns and a NPT (isothermal-isobaric) simulation of 400 ps respectively. Productive MD simulations were carried out in the NPT ensemble at 300 K, 1 atm with a time step of 2 fs. Molecular dynamics simulations covered a time of 500 ns for system containing Mre11 dimer and 300 ns for system containing Mre11 dimer and dsDNA molecule.

20.3 Root mean square deviation

The Root Mean Square Deviation (RMSD) of atomic positions is the measure of the average distance between the atoms (typically the backbone atoms) of two superimposed proteins. RMSD is computed as followed:

$$RMSD(v,w) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} ||v_i - w_i||^2}$$

where n is the number of backbone atoms and v_i and w_i are the coordinates of atom i of proteins v (target structure) and w (reference structure).

The RMSD is calculated between a defined starting point (reference structure) of the simulation and all succeeding frames (targets structure). This technique is widely used to determine if the molecular simulation reach the equilibrium state.

20.4 Root mean square fluctuation

The Root Mean Square Fluctuation (RMSF) is useful to calculate the fluctuation of an atom/residue with its mean in trajectory files, this fluctuation represents an estimate of the flexibility of protein structure over the simulation time. RMSF is computed as followed:

$$RMSF = \sqrt{\frac{1}{T}\sum_{t_j=1}^{T} (x_i(t_j) - \tilde{x}_i)^2}$$

where T is the duration of the simulation and $x_i(t_j)$ the coordinates of atom x_i at time t_j.

20.5 H-bonds

Hydrogen bonds are determined based on cutoffs for the angle Hydrogen - Donor - Acceptor (30°) and the distance donor – acceptor (3.5Å). OH and NH groups are regarded as donors, O and N are acceptors.

20.6 Protein secondary structure

Secondary structure analysis was performed by the defined secondary structure of proteins (DSSP) method[205] in order to highlight possible structural changes due to the substitution R10T.

20.7 Principal Component analysis

The identification of relevant motions in MD trajectories is carried out using the PCA that is able to filter global, collective and slow motions (relevant) from local and fast motions (not relevant). This operation is accomplished projecting the trajectories onto the eigenvectors to give the principal components. Eigenvectors are the columns of the diagonalized mass-weighted covariance matrix of the atomic positional fluctuations [168] which is calculated on protein backbone atoms. The element C_{ij} in the covariance matrix is given by

$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$$

where x_i (x_j) is the coordinate of the i^{th} (j^{th}) atom and $\langle \cdots \rangle$represents an ensemble average. The eigenvectors of the matrix describe the directions of the concerted motions while the eigenvalues give the magnitude of the motion for each direction. The first few principal components represent the most relevant motions and the first one contain the largest mean-square fluctuation. In this work PCA was performed with GROMACS 5.0.4 package.

20.8 Free energy landscape

To achieve a two-dimensional representation of FEL, we define the probability of finding the system in a given state α characterized by a value q_α of some quantity of interest as proportional to $e^{\frac{-G\alpha}{kT}}$ where G_α is the free energy of the state. FEL can be obtained from:

$$G_\alpha = -kT\ ln \left[\frac{P(q_\alpha)}{P_{max}(q)} \right]$$

where k is the Boltzmann constant, T is the temperature of simulation, $P(q_\alpha)$ is an estimate of the probability density function obtained from a histogram of the MD data and $P_{max}(q)$ is the probability of the most probable state. The two-dimensional free-energy landscapes are obtained from the joint probability of distributions $P(q_i, q_j)$ of the system where q_i and q_j are fluctuations along principal components.

20.9 Cluster analysis

The RMSD matrices are computed on trajectories of simulations without DNA by least square fitting on backbone atoms. These matrices are then processed using the Gromos [181] algorithms with a cutoff of 0.30 nm to extract clusters of similar conformations. The average structure (centroid) of clusters is defined as the protein structure with the lowest RMSD to all other structures belonging to the same cluster.

Chapter 21

Results and discussion

This work is in collaboration with research group of Professor Maria Pia Longhese. This group performs screens to identify mutants that suppress sensitivity of Sae2Δ mutants in order to understand the role of Sae protein in the activation of the endonuclease activity of Mre11 during the resection event. One of the most interesting mutant selected by Longhese's group is the Mre11-R10T. This mutant not only bypasses Sae2 function in DSB resection, but also resects the DSB more efficiently than wild type Mre11, indicating that this Mre11 mutant variant accelerates the resection process. Due to the fact that suppression of Sae2Δ by Mre11-R10T is independent of Mre11 nuclease activity (H125N substitution affects the Mre11 nuclease activity, the double mutant Mre11-R10T-H125N shows equivalent suppression of the sae2Δ DNA damage sensitivity) and that the lack of Exo1 dramatically reduced the efficiency of DSB resection in mre11-R10T cells, the increase of the DNA ends processing seems to be imputed to Exo1. The hypothesis proposed by Professor Longhese is that the substitution R10T in Mre11 protein can enhance Exo1 resection activity. Scope of this investigation is to identify the molecular determinants of Mre11-R10T using MD simulations that can explain the up-regulation of Exo1.

21.1 MD simulation of dimers

A model for *S. cerevisiae* Mre11 was obtained by homology modeling based on the structure of *S. pombe* homolog (see Materials and Methods for details). Based on the model, the R10 residue is localized on the β1 strand of the β-sheet within the endonuclease domain (Figure 21.1). In particular, R10 binds through a salt bridge and a hydrogen bond to D284 and through a hydrogen bond to D49, localized on different strands (respectively β7 and β2) of the same β-sheet, thus probably stabilizing the secondary structure (Figure 21.1). Moreover, R10 binds through a hydrogen bond to N46, localized on the loop at the end of the α1 helix (Figure 21.1).

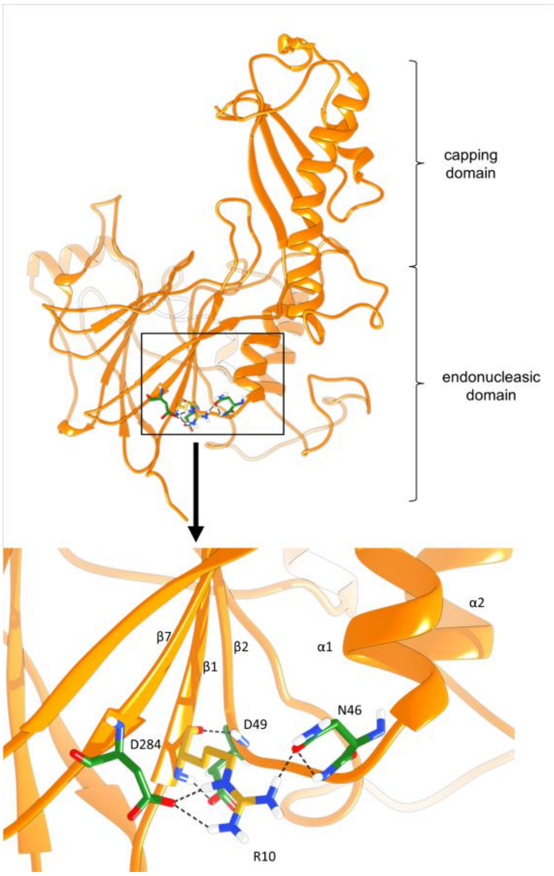


Figure 21.1: The position of residue R10 is shown in a model of monomeric Mre11 spanning endonuclease and capping domain. Residue R10 is shown in yellow. Residues interacting with R10 are shown in green.

Molecular dynamics simulations were performed on both Mre11 and Mre11-R10T dimers to investigate the effects of the substitution R10T at the molecular detail. The analyses of the MD trajectories have been carried out discarding the portions of the simulations required by the systems to reach stable values of backbone rmsd (about 50 ns) in order to be sure that calculated parameters reflect the intrinsic properties of the systems (Figure21.2-**a**). The conformational flexibility of Mre11 and Mre11-R10T has been investigated using the RMSF and monitoring the variation of secondary structure. RMSF of backbone atoms coordinates from their starting positions was calculated as a function of the residue number. As shown in Figure 21.2-**b** and 21.2-**c** the flexibility of the two dimers is practically the same except for residues Leu_{87}, Gln_{95}, Asp_{351} and Asn_{376} of subunit A (21.2-**b**) that are more rigid ($\simeq$ 0.2 nm) in wild type than in mutant. The aforementioned residues are not in close proximity with Arg_{10} thus it is not trivial to understand how the substitution R10T can increase their mobility, furthermore the region around the R10 residue does not seem to be affected by the mutation in term of mobility. Overall, the secondary structure of the two dimers is stable during the simulations and the are no differences between Mre11 and Mre11-R10T (Figure 21.3). Principal component analysis (PCA) was used to obtain the essential motions of the systems during the simulations. The analysis has been focused on the first three eigenvectors that represent more than 60% of total variance (Figure 21.4) of the motion for either Mre11 dimers. The motion described by the three selected eigenvectors is essentially the rotation of the capping domains (Figure 21.4). The three eigenvectors are used to evaluate the free-energy landscapes (FELs) based on the first three principal components (PCs): the bi-dimensional projections of PC1 vs. PC2 and PC1 vs. PC3 have been analyzed. The FELs of Mre11 feature only one minimum energy basin, whereas Mre11_R10T FELs show two minimum energy basins (Figure 21.5). Cluster analysis (see Appendix C) combined to FEL description was used to establish whether conformations that belong to the same energy basin are similar. Results show that the WT Mre11 dimer has only one stable structure that belongs to cluster-1, whereas, in presence of the substitution R10T the dimer can exist in two different conformations that are represented by cluster-1 and cluster-2 structures (Figure 21.5). The pairwise superposition of the aforementioned structures pointed out an important difference between the native and mutant Mre11 systems: the Mre11-A subunit of the mutant dimer shows an accentuated rotation of the capping domain respect to the wild type counterpart (Figure 21.6-21.7).

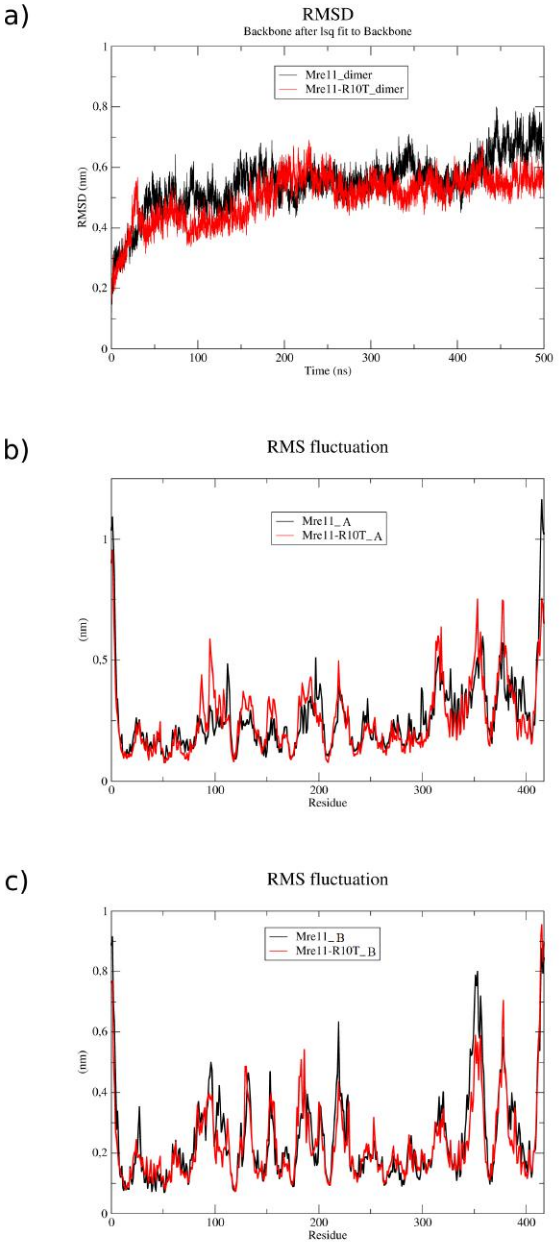


Figure 21.2: Figure reports beckbone RMSD for Mre11 and Mre11-R10T dimers **(a)**; average RMSF of Mre11 and Mre11-R10T monomer A **(b)**; average RMSF of Mre11 and Mre11-R10T monomer A **(c)**. Wild type Mre11 is reported in black while mutant Mre-R10T in red.

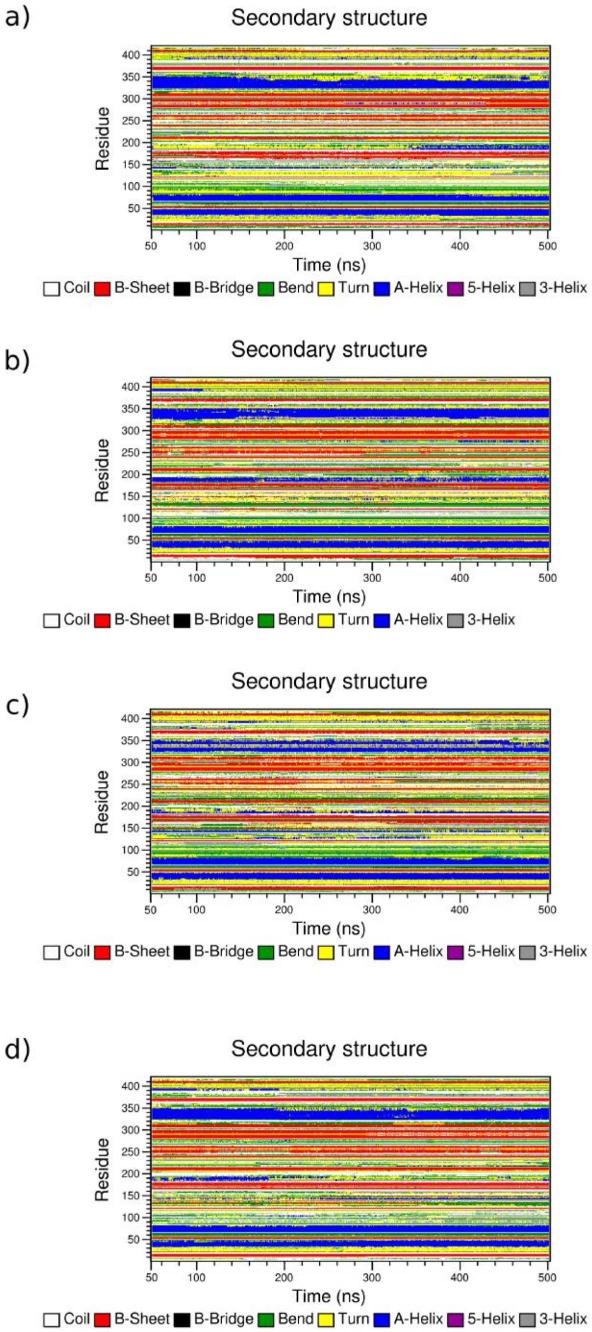


Figure 21.3: Time evolution of the secondary structure content of Mre11_subunitA **(a)**, Mre11_subunitB **(b)**, Mre11-R10T_subunitA **(c)**, Mre11-R10T_subunitA **(d)**.

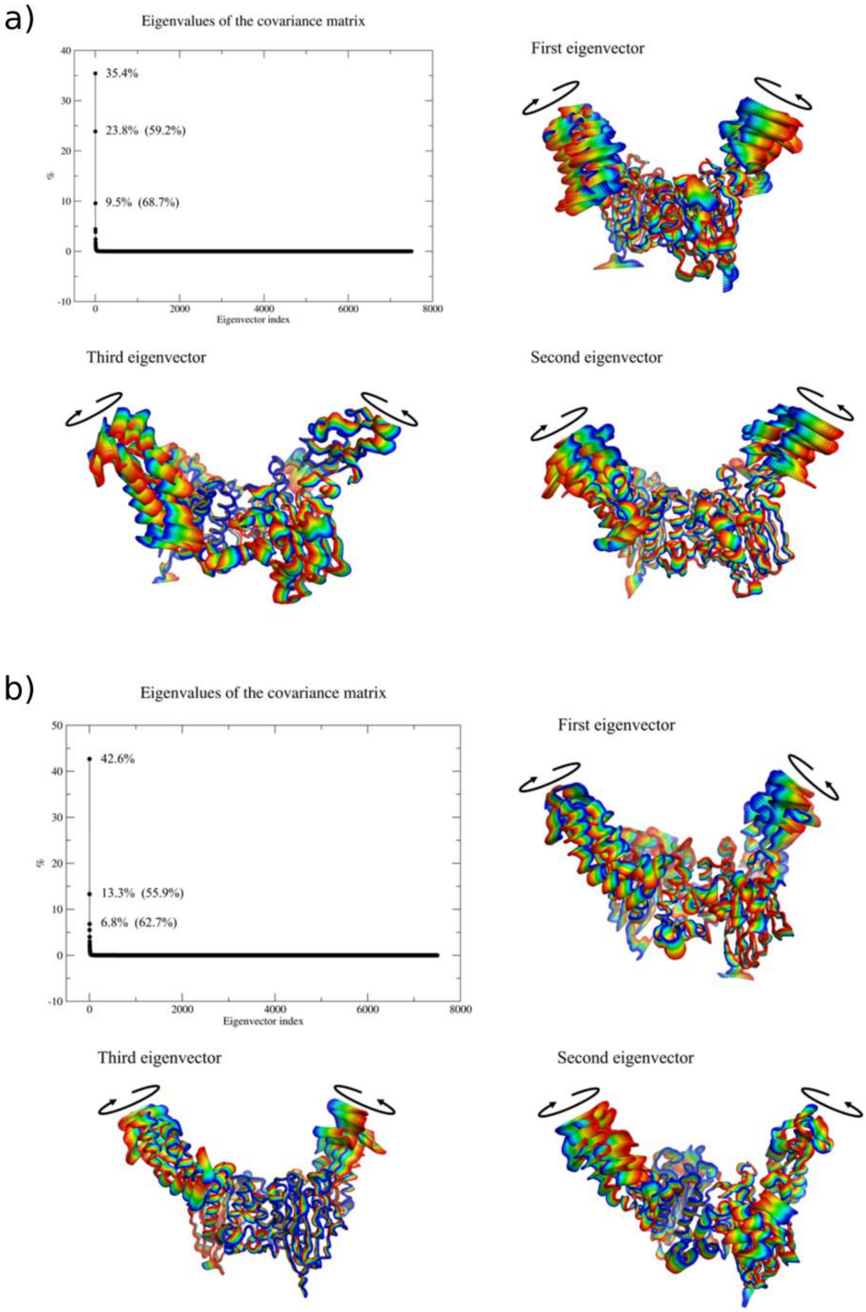


Figure 21.4: Percentage of distribution of total variance across the eigenvector for wild type Mre11 **(a)** and mutant Mre11-R10T **(b)**. It has been shown the percentage variance, the cumulative variance (value in brackets) and the motions for the first three eigenvectors. The motion along the three eigenvectors is described by a static representation indicating the starting point in blue and the final point in red.

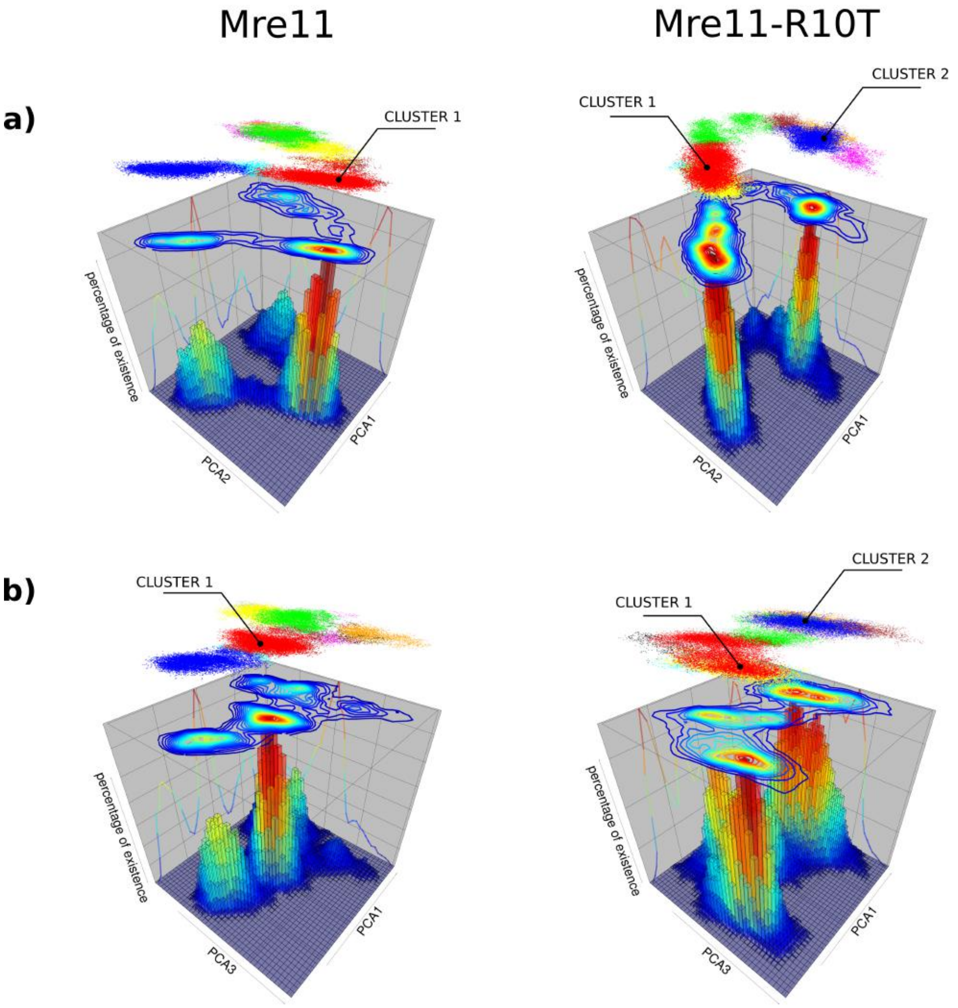


Figure 21.5: Free-energy landscape of Mre11 and Mre11-R10T. FEL (contour plot) and percentage of existence (3D-histogram) using as reaction coordinates the projection of trajectory along PC1-PC2 **(a)** and PC1-PC3 **(b)**. Clustering results of Gromos algorithm are reported above each FELs.

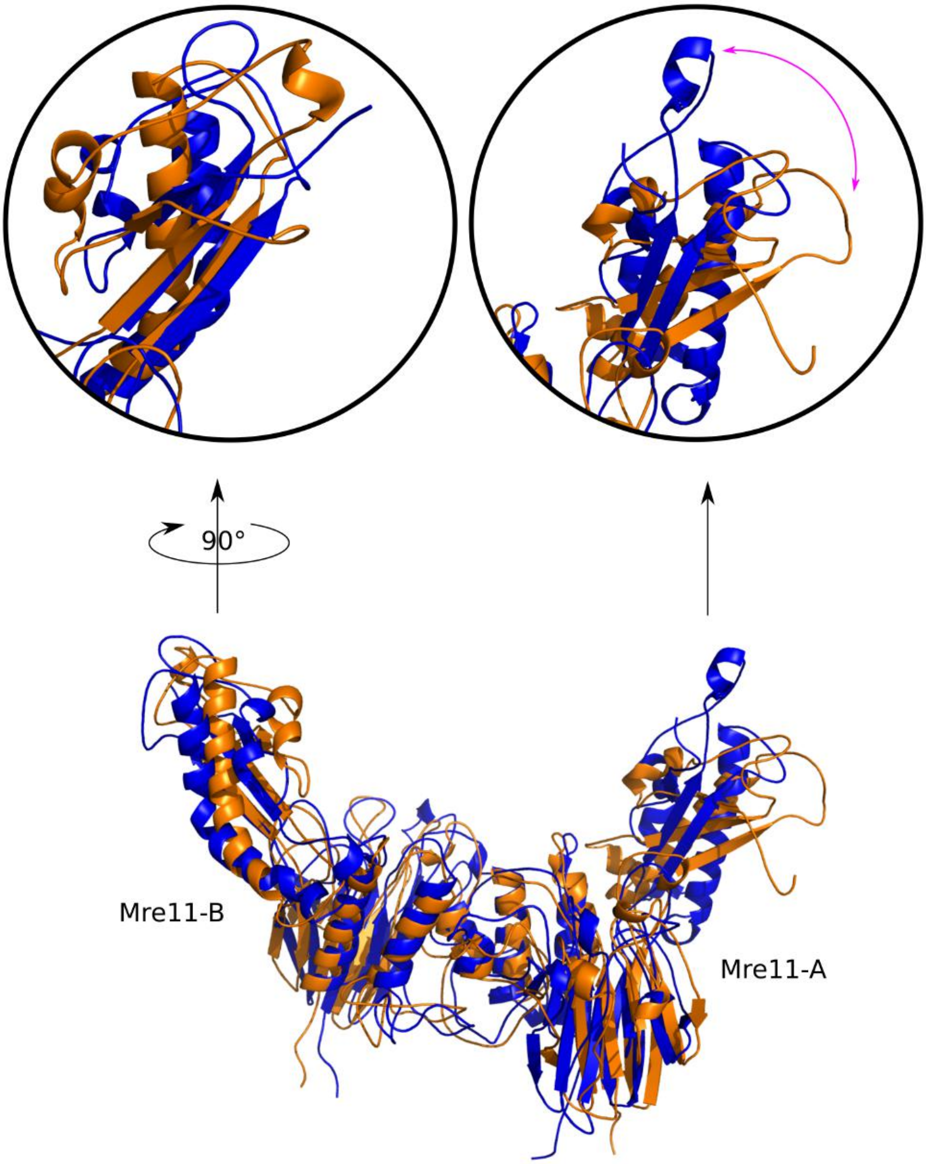


Figure 21.6: Superposition of centroid structures of cluster-1-Mre11 (orange) and cluster-1-Mre11-R10T (blue). The blown-up pictures of the capping domain of chains A and B are shown on the right and on the left respectively.

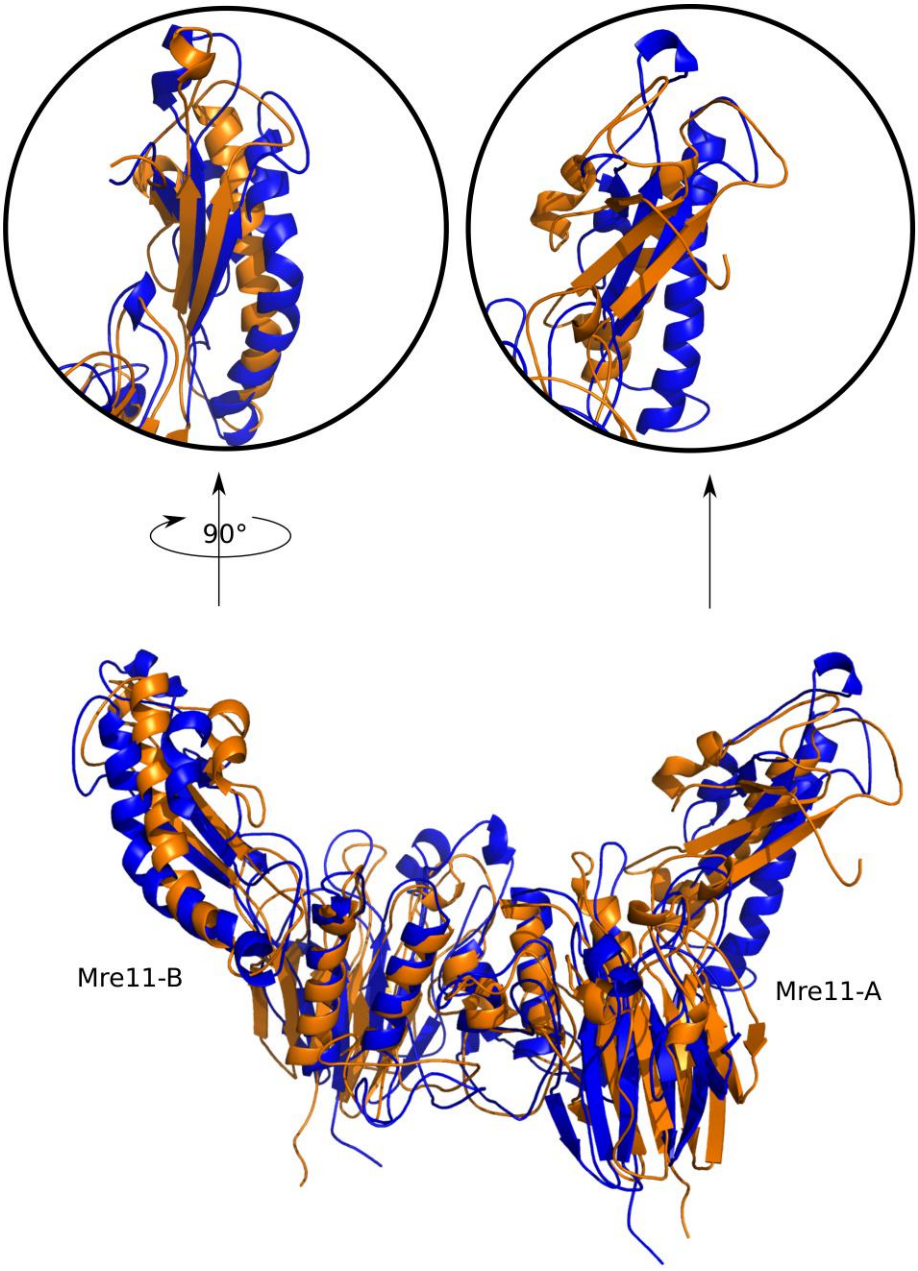


Figure 21.7: Superposition of centroid structures of cluster-1-Mre11 (orange) and cluster-2-Mre11-R10T (blue). The blown-up pictures of the capping domain of chains A and B are shown on the right and on the left respectively.

21.2 DNA binding

Since the capping domain is known to be involved in interaction with DNA, the possible effects of this abnormal rotation in the mutant protein have been further investigated by performing a new MD simulation in presence of a dsDNA fragment (14 nucleotide pairs). Starting points of the simulations were prepared using the average structures from cluster 1 of both Mre11 dimers and the dsDNA was located in close proximity of the dimerization interface according to the crystal structure deposited in the PDB (code: 4TUG; Figure 21.10-**a**). The MD production phase was reached after 25 ns as shown from RMSD analysis for both systems (Figure 21.8). During the simulations (both trajectories are filtered along the 1^{st} eigenvector) the dsDNA promptly rearranged from the initial orientation, roughly according to a 90° rotation about an axis which is orthogonal to the longitudinal axis of the double helix. Such rearrangement causes the DNA-ends to encounter rapidly the two capping domains of the two Mre11 subunits (Figure **21.10-b**). After the capture of DNA-ends by the capping domains, the DNA double helix is guided into the the nuclease site in both complexes. Interaction between DNA and Mre11 was monitored evaluating the number of hydrogen bonds formed between DNA and dimers. Both systems follow a similar pattern of H-bonds (Figure 21.9), this fact indicates that the substitution does not affect the DNA binding with Mre11. Although both WT Mre11 and Mre11-R10T show a similar DNA unwinding activity as shown by the increase of the length of helix turn, the Mre11-R10T shows also the displacement of the DNA double helix (Figure **21.10-c**).This event appears to be related to the “super” rotation of the capping domain of one Mre11 monomer, such that the motion of this misaligned capping domain allows it to wedge in between the two DNA strands, causing the melting of the dsDNA terminus.

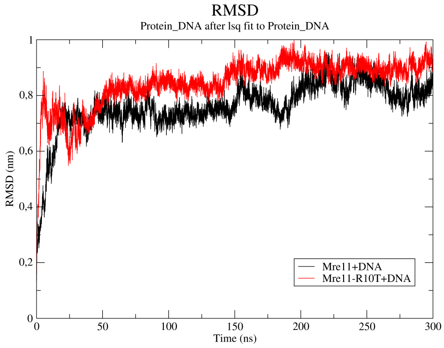


Figure 21.8: Figure reports beckbone RMSD for Mre11+DNA and Mre11-R10T+DNA dimers. Wild type system is reported in black while mutant system in red.

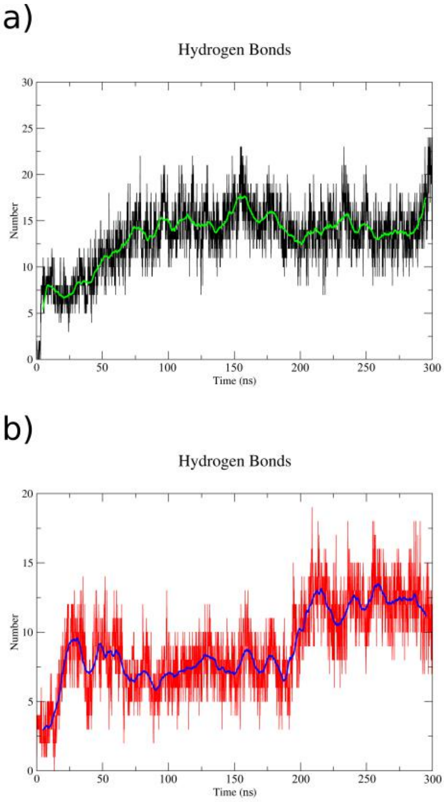


Figure 21.9: Time evolution of the number of intermolecular hydrogen bonds formed between dsDNA and protein complex: system Mre11+DNA is reported in black (**a**), system Mre11-R10T+DNA is reported in red (**b**). Green line and blue line represent the running average with a length of 100.

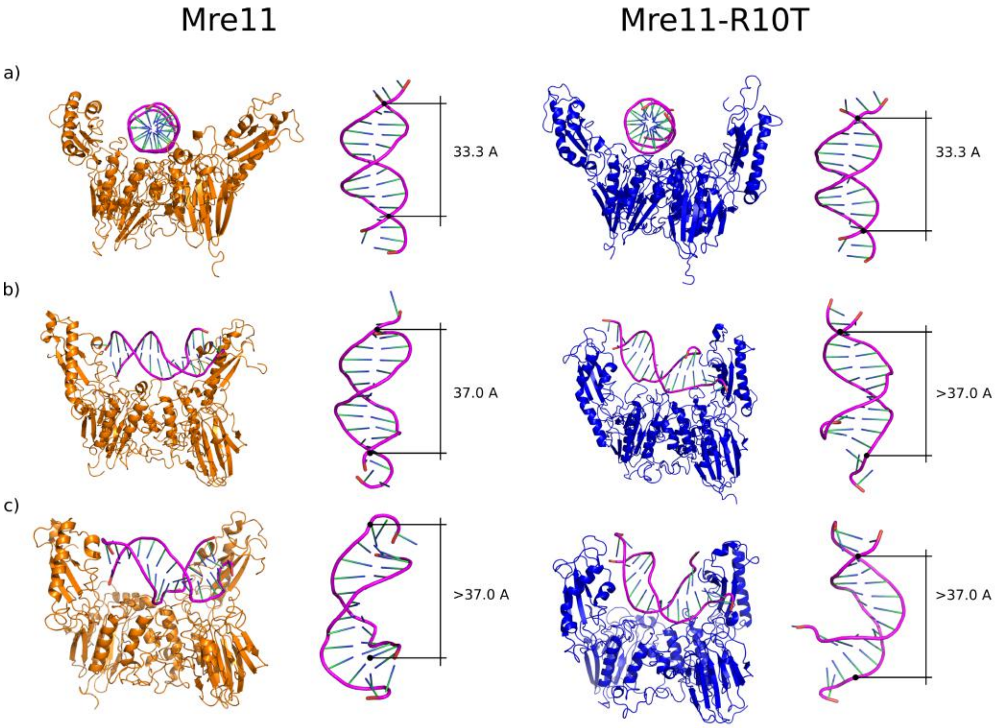


Figure 21.10: Snapshots of the key steps of the MD trajectories of Mre11 and Mre11-R10T filtered along the first eigenvector: in the starting point the DNA orientation corresponds to that of the crystallographic structure (4TUG) **(a)**, interaction mode of the dsDNA fragment with the capping domains **(b)** and interaction of the DNA with endonuclease site of Mre11 dimer **(c)**. For each step it has been reported the change of length of helix turn (distance is reported Ȧ)

Chapter 22

Conclusions

By molecular dynamics simulations on wild type Mre11 dimer with dsDNA, it is found that the DNA ends take contact rapidly with the two Mre11 capping domains and such positioning causes a strain and partial unwinding of the DNA double helix. Since Exo1 was shown to possess a higher affinity for branched DNA end structures relative to linear dsDNA ends [195], this MRX-mediated DNA unwinding might generate a higher-affinity binding of Exo1 to the DSB end. When a similar analysis was performed on the Mre11-R10T dimer, one Mre11-R10T subunit undergoes an abnormal rotation of the capping domain with respect to the wild type Mre11 counterpart. This altered rotation causes the capping domain to wedge in between the two DNA strands, leading to the opening of the DNA double helix holding the branched structure open. This analysis suggests that Mre11-R10T enhances Exo1-mediated DSB resection by causing a persistent melting of the dsDNA that provides to Exo1 a better substrate to bind and resect. In conclusion, MRX can directly control the efficiency of Exo1-mediated DSB resection independently of Mre11 nuclease activity by promoting a change in DNA topology.

Bibliography

[167] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, sep 2015.

[168] Andrea Amadei, Antonius B. M. Linssen, and Herman J. C. Berendsen. Essential dynamics of proteins. *Proteins: Structure, Function, and Genetics*, 17(4):412–425, dec 1993.

[169] J A Y L Banks, Hege S Beard, Yixiang Cao, A R T E Cho, Wolfgang Damm, Ramy Farid, Anthony K Felts, Thomas A Halgren, Daniel T Mainz, J O N R, Robert Murphy, Dean M Philipp, Matthew P Repasky, Linda Y, Bruce J Berne, Richard A Friesner, Emilio Gallicchio, and M Levy. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *Journal of computational chemistry*, 26(16):1752–1780, 2005.

[170] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *The Journal of Physical Chemistry*, 1987.

[171] P R Blier, A J Griffith, J Craft, and J A Hardin. Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. *The Journal of biological chemistry*, 268(10):7594–601, apr 1993.

[172] E. Cannavo, P. Cejka, and S. C. Kowalczykowski. Relationship of DNA degradation by Saccharomyces cerevisiae Exonuclease 1 and its stimulation by RPA and Mre11-Rad50-Xrs2 to DNA end resection. *Proceedings of the National Academy of Sciences*, 110(18):E1661–E1668, apr 2013.

[173] Elda Cannavo and Petr Cejka. Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. *Nature*, 514(7520):122–125, sep 2014.

[174] Petr Cejka, Elda Cannavo, Piotr Polaczek, Taro Masuda-Sasa, Subhash Pokharel, Judith L. Campbell, and Stephen C. Kowalczykowski. DNA end resection by Dna2 Sgs1 RPA and its stimulation by Top3 Rmi1 and Mre11 Rad50 Xrs2. *Nature*, 467(7311):112–116, sep 2010.

[175] Pauline Chanut, Sébastien Britton, Julia Coates, Stephen P Jackson, and Patrick Calsou. Coordinated nuclease activities counteract Ku at single-ended DNA double-strand breaks. *Nature communications*, 7:12889, sep 2016.

[176] Swaroop Chatterjee, Pablo G. Debenedetti, Frank H. Stillinger, and Ruth M. Lynden-Bell. A computational investigation of thermodynamics, structure, dynamics and solvation behavior in modified water models. *The Journal of Chemical Physics*, 128(12):124511, mar 2008.

[177] L Chen, K Trujillo, W Ramos, P Sung, and A E Tomkinson. Promotion of Dnl4-catalyzed DNA end-joining by the Rad50-Mre11-Xrs2 and Hdf1-Hdf2 complexes. *Molecular cell*, 8(5):1105–15, nov 2001.

[178] K. K. Chiruvella, Z. Liang, and T. E. Wilson. Repair of Double-Strand Breaks by End Joining. *Cold Spring Harbor Perspectives in Biology*, 5(5):a012757–a012757, may 2013.

[179] Michela Clerici, Davide Mantiero, Ilaria Guerini, Giovanna Lucchini, and Maria Pia Longhese. The Yku70-Yku80 complex contributes to regulate double-strand break processing and checkpoint activation during the cell cycle. *EMBO reports*, 9(8):810–8, aug 2008.

[180] Michela Clerici, Davide Mantiero, Giovanna Lucchini, and Maria Pia Longhese. The Saccharomyces cerevisiae Sae2 protein promotes resection and bridging of double strand break ends. *The Journal of biological chemistry*, 280(46):38631–8, nov 2005.

[181] Xavier Daura, Karl Gademann, Bernhard Jaun, Dieter Seebach, Wilfred F Van Gunsteren, and Alan E Mark. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed*, 38(12), 1999.

[182] W S Dynan and S Yoo. Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids. *Nucleic acids research*, 26(7):1551–9, apr 1998.

[183] S. S. Foster, A. Balestrini, and J. H. J. Petrini. Functional Interplay of the Mre11 Nuclease and Ku in the Response to Replication-Associated DNA Damage. *Molecular and Cellular Biology*, 31(21):4379–4389, nov 2011.

[184] A J Griffith, J Craft, J Evans, T Mimori, and J A Hardin. Nucleotide sequence and genomic structure analyses of the p70 subunit of the human Ku autoantigen: evidence for a family of genes encoding Ku (p70)-related polypeptides. *Molecular biology reports*, 16(2):91–7, may 1992.

[185] Berendsen H.J.C., D. van der Spoel, and R.van Drunen. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56, sep 1995.

[186] Matthew P. Jacobson, David L. Pincus, Chaya S. Rapp, Tyler J.F. Day, Barry Honig, David E. Shaw, and Richard A. Friesner. A hierarchical approach to all-atom protein loop prediction. *Proteins: Structure, Function, and Bioinformatics*, 55(2):351–367, mar 2004.

[187] M.P. Jacobson, R.A. Friesner, Z. Xiang, and B. Honig. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *Journal of Molecular Biology*, 320(3):597–608, jul 2002.

[188] Hui-I Kao, Judith L Campbell, and Robert A Bambara. Dna2p helicase-nuclease is a tracking protein, like FEN1, for flap cleavage during Okazaki fragment maturation. *The Journal of biological chemistry*, 279(49):50840–9, dec 2004.

[189] S E Lee, J K Moore, A Holmes, K Umezu, R D Kolodner, and J E Haber. Saccharomyces Ku70, mre11-rad50 and RPA proteins regulate adaptation to G2/M arrest after DNA damage. *Cell*, 94(3):399–409, aug 1998.

[190] Michael Lisby, Jacqueline H. Barlow, Rebecca C. Burgess, and Rodney Rothstein. Choreography of the DNA Damage Response. *Cell*, 118(6):699–713, sep 2004.

[191] A. Mehta and J. E. Haber. Sources of DNA Double-Strand Breaks and Models of Recombinational DNA Repair. *Cold Spring Harbor Perspectives in Biology*, 6(9):a016428–a016428, sep 2014.

[192] Eleni P. Mimitou and Lorraine S. Symington. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature*, 455(7214):770–774, oct 2008.

[193] Eleni P Mimitou and Lorraine S Symington. Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. *The EMBO Journal*, 29(19):3358–3369, oct 2010.

[194] T Mimori and J A Hardin. Mechanism of interaction between Ku protein and DNA. *The Journal of biological chemistry*, 261(22):10375–9, aug 1986.

[195] Matthew L Nicolette, Kihoon Lee, Zhi Guo, Mridula Rani, Julia M Chow, Sang Eun Lee, and Tanya T Paull. Mre11-Rad50-Xrs2 and Sae2 promote 5' strand resection of DNA double strand breaks. *Nature Structural & Molecular Biology*, 17(12):1478–1485, dec 2010.

[196] A. V. Nimonkar, J. Genschel, E. Kinoshita, P. Polaczek, J. L. Campbell, C. Wyman, P. Modrich, and S. C. Kowalczykowski. BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes & Development*, 25(4):350–362, feb 2011.

[197] Hengyao Niu, Woo-Hyun Chung, Zhu Zhu, Youngho Kwon, Weixing Zhao, Peter Chi, Rohit Prakash, Changhyun Seong, Dongqing Liu, Lucy Lu, Grzegorz Ira, and Patrick Sung. Mechanism of the ATP-dependent DNA end-resection machinery from Saccharomyces cerevisiae. *Nature*, 467(7311):108–111, sep 2010.

[198] Eun Yong Shim, Woo-Hyun Chung, Matthew L Nicolette, Yu Zhang, Melody Davis, Zhu Zhu, Tanya T Paull, Grzegorz Ira, and Sang Eun Lee. Saccharomyces cerevisiae Mre11-Rad50-Xrs2 and Ku proteins regulate association of Exo1 and Dna2 with DNA breaks. *The EMBO Journal*, 29(19):3370–3380, oct 2010.

[199] Lorraine S. Symington. Mechanism and regulation of DNA end resection in eukaryotes. *Critical Reviews in Biochemistry and Molecular Biology*, 51(3):195–212, may 2016.

[200] Phuoc T Tran, Naz Erdeniz, Sandra Dudley, and R Michael Liskay. Characterization of nuclease-dependent functions of Exo1p in Saccharomyces cerevisiae. *DNA repair*, 1(11):895–912, nov 2002.

[201] Matteo Villa, Corinne Cassani, Elisa Gobbini, Diego Bonetti, and Maria Pia Longhese. Coupling end resection with the checkpoint response at DNA double-strand breaks. *Cellular and Molecular Life Sciences*, 73(19):3655–3663, oct 2016.

[202] John R. Walker, Richard A. Corpina, and Jonathan Goldberg. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature*, 412(6847):607–614, aug 2001.

[203] Junmei Wang, Piotr Cieplak, Peter A Kollman, and P A Kollman. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *Journal of Computational Chemistry J Comput Chem*, 21(21):1049–1074, 2000.

[204] D. Wu, L. M. Topper, and T. E. Wilson. Recruitment and Dissociation of Nonhomologous End Joining Proteins at a DNA Double-Strand Break in Saccharomyces cerevisiae. *Genetics*, 178(3):1237–1249, feb 2008.

[205] Yun Xie, Jian Zhou, and Shaoyi Jiang. Parallel tempering Monte Carlo simulations of lysozyme orientation on charged surfaces. *The Journal of Chemical Physics*, 132(6):065101, feb 2010.

[206] Yu Zhang, Melissa L Hefferin, Ling Chen, Eun Yong Shim, Hui-Min Tseng, Youngho Kwon, Patrick Sung, Sang Eun Lee, and Alan E Tomkinson. Role of Dnl4-Lif1 in nonhomologous end-joining repair complex assembly and suppression of homologous recombination. *Nature Structural & Molecular Biology*, 14(7):639–646, jul 2007.

[207] Zhu Zhu, Woo-Hyun Chung, Eun Yong Shim, Sang Eun Lee, and Grzegorz Ira. Sgs1 Helicase and Two Nucleases Dna2 and Exo1 Resect DNA Double-Strand Break Ends. *Cell*, 134(6):981–994, sep 2008.

Chapter 23

Conclusions and remarks

In this dissertation different computational approaches, spanning from QM to MM methods, have been used to investigate and disclose some phenomena at molecular level, proving to be powerful and useful tools to describe different relevant aspects of biologically related macromolecules. Such aspects include the characterization of molecular structure (and of the related electronic structure), catalytic reactivity and dynamic properties of the investigated biomolecules, all representing an important subject of research in biotechnology. It is increasingly evident that computational methods not only provide fast and cheap strategies to address specific issues that generally covers all research areas that are focused on the molecular detail of phenomena (may these physical, chemical or biological) but they often represent the only way to gain detailed insights from the atomic point of view. Computational quantum chemistry is a solid tool to thoroughly investigate the electron properties, the 3-D structure and the catalytic pathways of enzymes as illustrated in the present work.

DFT simulations have been used to characterize the structure of FeMo-co at the E_4 state. Results show two iron-bridging hydrides located on Fe_2 and Fe_6 in agreement with EPR data. This is a clear example of how the computational techniques can assist the experimental techniques to gain a more detailed information. EPR spectroscopy can provide insights about the stereochemical arrangements of the hydride ligand (in this case EPR predicts a Fe-H-Fe hydrides) but it is not able to show their exact location within the enzyme cofactor. Moreover, DFT simulations have allowed to identify the most probable oxidation state among those proposed in literature for iron atoms of the cofactor. It is worth to noting that the redox state of atoms is not a real physically observable-quantity and it can be estimated only through computational methods. The proposed oxidation state of FeMo-co Fe's is the configuration [4Fe^{II}:3Fe^{III}] with a (3↑:4↓) spin coupling in presence of a protonated homocitrate.

DFT has been used also to refine the X-ray structure of $96^{Arg\rightarrow Gln}$ Mo-nitrogenase variant of *A. vinelandii.* X-ray structure of this kind of nitrogenase reveals an elongated electron density consistent with the presence of acetylene molecule nearby the active face of FeMo-co. Computational simulations confirm the presence of the acetylene inside the mutant active site on energetical basis and furthermore they provide an explanation of why this variant is able to accommodation acetylene close to FeMo-co. Substitution $96^{Arg\rightarrow Gln}$ alters the chemico-physical properties of the active site increasing the hydrophobic nature of the pocket. Due to the non-polar nature of the acetylene, the mutant pocket is more suitable for hosting such substrate.

DFT has also revealed a useful tool to study and to characterize the catalytic pathway of a *de novo* designed carbonic anhydrase. The techniques leading to *de novo* designed enzymes represent a final goal of biotechnologies because they allow to design new proteins with new catalytic activities which can be used at the industrial level. Here DFT has been used to disclose the intimate reaction mechanism of the ester hydrolysis catalyzed by a synthetic model of carbonic anhydrase with the hope of improving the efficiency of the natural enzyme. Results indicate that the rate determining step of the entire catalytic process is the release of the products hydrolisis (specifically the acid moiety) from the catalyst. Moreover the effect of a mutation has been tested *in silico.* In the natural carbonic anhydrase a key residue implicated in the reaction is a threonine. For such reason it has been implemented the substitution $19^{Lue\rightarrow Thr}$ with the aim to enhance the catalytic rate of the enzyme. Notably, both experimental and computational data reveal that this mutation leads to an unexpected decrease of catalytic activity. By DFT, it has been demonstrated that the Thr residue within the synthetic peptide induces the increase of the pk$_a$ of Zn-H$_2$O. Increments of pk$_a$ disfavor the formation of the activate water (Zn-OH$^-$) that is essential to the catalysis.

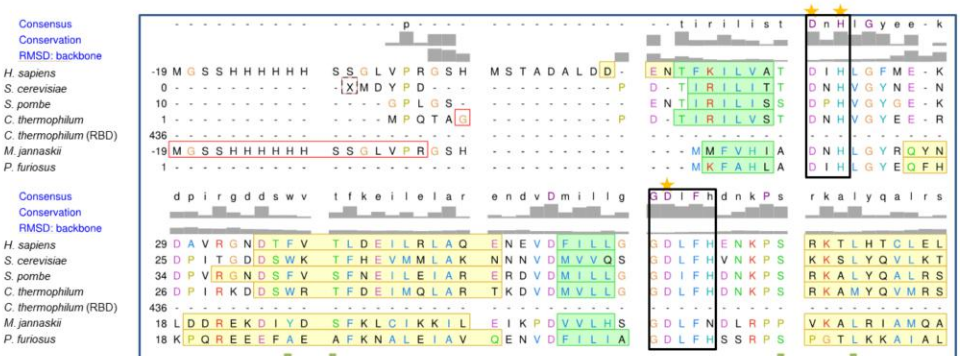
The last project of this dissertation illustrates the characterization of different behaviors of Mre11 and mutated Mre11-R10T, a subunit of Mre11-Rad50-Xrs2 complex involved in DNA double strand break repair, using Force Field approaches (Molecular

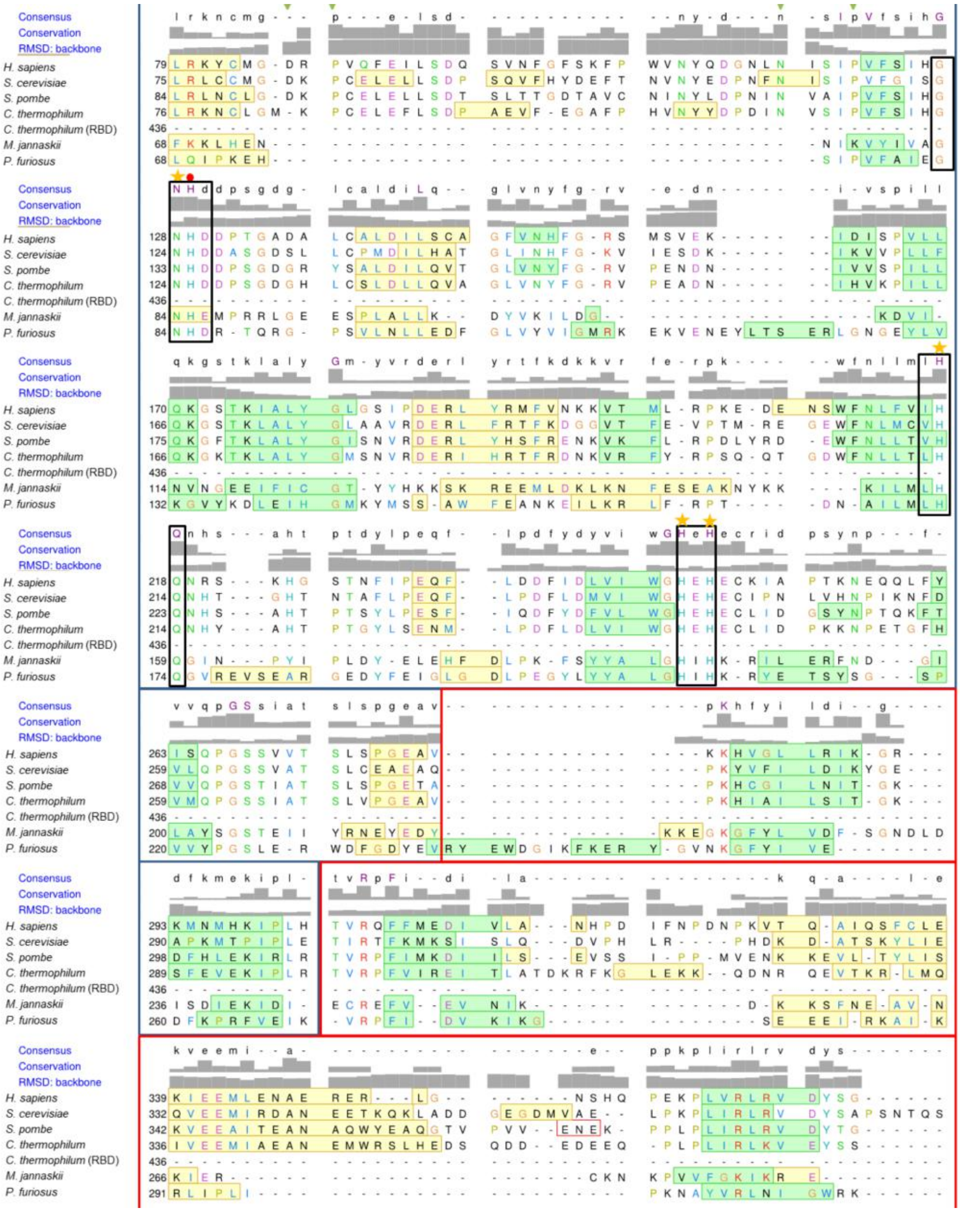
Mechanics and Dynamics, both based on the classical physics in contrast to DFT technique, based on quantum mechanics). MD simulations show a viable interaction mode of Mre11 and double strand DNA. Capping domains of Mre11 dimer take contact with DNA ends and they guide the double helix into the nuclease site of Mre11. Moreover, while the interaction takes place, capping domains cause a strain and partial unwinding of the dsDNA. In Mre11-R10T such domains undergo an anomalous rotation leading to the opening of the DNA double helix holding the branched structure open. Only with molecular dynamics approach it has been possible to gain insights into the molecular details of this dynamic rearrangement that occurs within cells and that involves complicated interactions between enzyme and DNA.

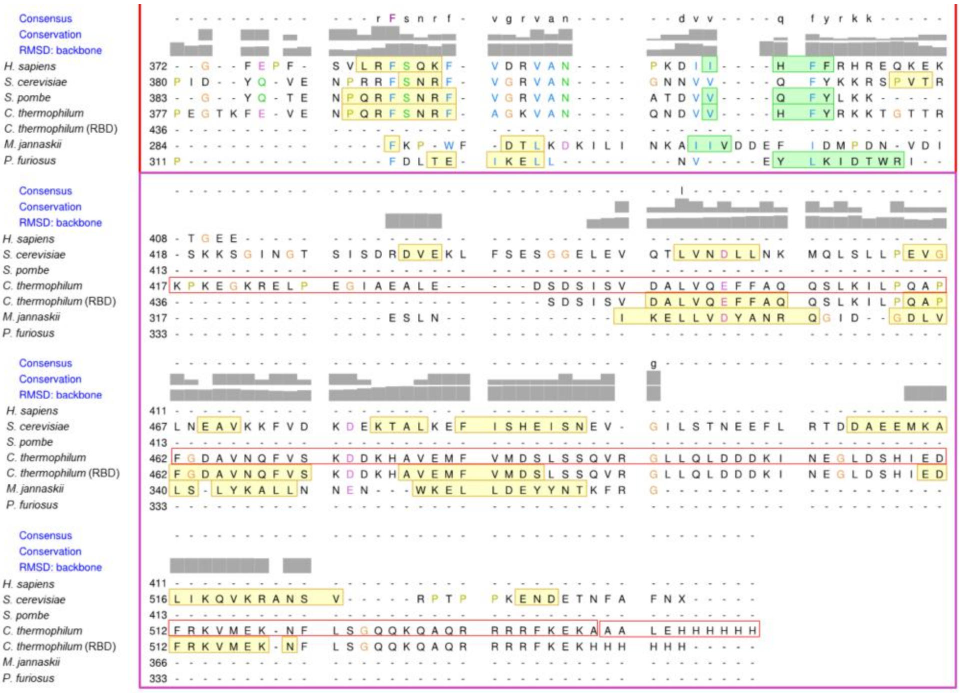
Appendix

Appendix A

Structural alignment of Mre11 orthologs from Homo sapiens (PDB 3T1I), *Schizosaccharomyces cerevisiae* (PDB 4FCX), *Chaetomium thermophylum* (PDB 4KYE and 5DA9 for RBD region), *Methanocaldococcus jannaskii* (PDB 3AV0), *Pyrococcus furiosus* (PDB 1II7). Secondary structures are indicated, wherever the structure is available (otherwise the sequence is marked by a red rectangle), as follows: yellow rectangles, alpha helix; green rectangles, beta sheets. Alignment residues coloring is according to ClustalW style. The consensus sequence, and histograms representing residue conservation and backbone RMSD are indicated on top. Functional regions of the proteins are indicated by open boxes: blue box, endonucleasic domain; red box, capping domain; purple box, Rad50-binding domain (RBD); orange box, latching loop. Black open boxes highlight endonucleas motifs. Stars are represented above the consensus for Mg^{2+} binding residues, relevant for endonucleasic activity; a red dot marks the catalytic His residue; green triangles indicate the residues identified for Xrs2/Nbs1 binding in SpMre11

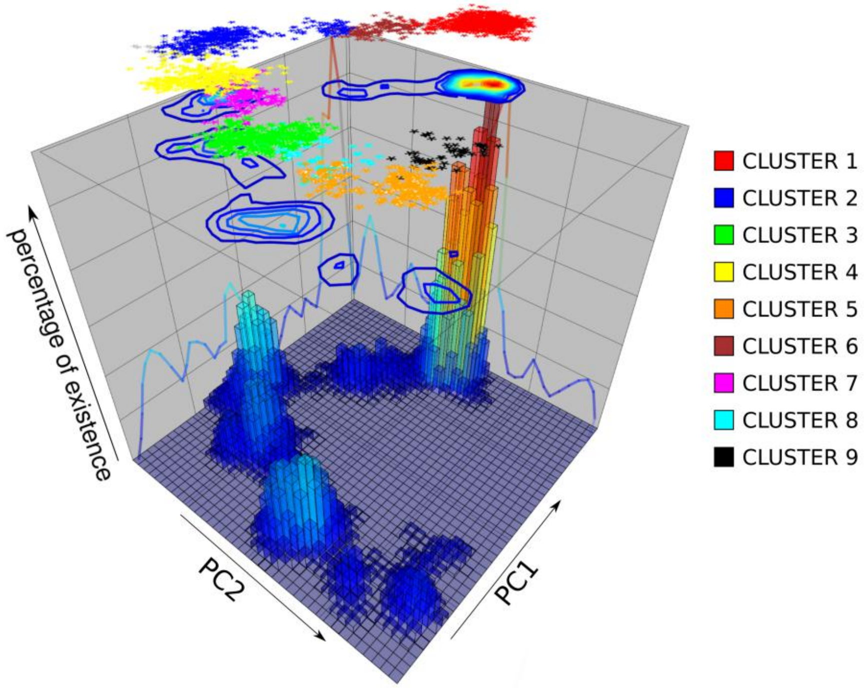



Consensus l r k n c m g - - - p - - - e - l s d - - - - - - - - - - - - - - n y - d - - - n - s l p V f s i h G
Conservation
RMSD: backbone
H. sapiens 79 LRKYCMG-DR PVQFEILSDQ SVNFGFSKFP WVNYQDGNLN ISIPVFSIHG
S. cerevisiae 75 LRLCCMG-DK PCELELLSDP SQVFHYDEFT NVNYEDPNFN ISIPVFGISG
S. pombe 84 LRLNCLG-DK PCELELLSDT SLTTGDTAVC NINYLDPNIN VAIPVFSIHG
C. thermophilum 76 LRKNCLGM-K PCELEFLSDP AEVF-EGAFP HVNYYDPDIN VSIPVFSIHG
C. thermophilum (RBD) 436 ---------- ---------- ---------- ---------- ----------
M. jannaskii 68 FKKLHEN--- ---------- ---------- ---------- -NIKVYIVAG
P. furiosus 68 LQIPKEH--- ---------- ---------- ---------- -SIPVFAIEG
Consensus N H d d p s g d g - l c a l d i L q - - g l v n y f g - r v - e - d n - - - - - - - i - v s p i l l
Conservation
RMSD: backbone
H. sapiens 128 NHDDPTGADA LCALDILSCA GFVNHFG-RS MSVEK----- --IDISPVLL
S. cerevisiae 124 NHDDASGDSL LCPMDILHAT GLINHFG-KV IESDK----- --IKVVPLLF
S. pombe 133 NHDDPSGDGR YSALDILQVT GLVNYFG-RV PENDN----- --IVVSPILL
C. thermophilum 124 NHDDPSGDGH LCSLDLLQVA GLVNYFG-RV PEADN----- --IHVKPILL
C. thermophilum (RBD) 436 ---------- ---------- ---------- ---------- ----------
M. jannaskii 84 NHEMPRRLGE ESPLALLK-- DYVKILDG-- ---------- ------KDVI-
P. furiosus 84 NHDR-TQRG- PSVLNLLEDF GLVYVIGMRK EKVENEYLTS ERLGNGEYLV
Consensus q k g s t k l a l y G m - y v r d e r l y r t f k d k k v r f e - r p k - - - - - - w f n l l m l H
Conservation
RMSD: backbone
H. sapiens 170 QKGSTKIALY GLGSIPDERL YRMFVNKKVT ML-RPKE-DE NSWFNLFVIH
S. cerevisiae 166 QKGSTKLALY GLAAVRDERL FRTFKDGGVT FE-VPTM-RE GEWFNLMCVH
S. pombe 175 QKGFTKLALY GISNVRDERL YHSFRENKVK FL-RPDLYRD -EWFNLLTVH
C. thermophilum 166 QKGKTKLALY GMSNVRDERI HRTFRDNKVR FY-RPSQ-QT GDWFNLLTLH
C. thermophilum (RBD) 436 ---------- ---------- ---------- ---------- ----------
M. jannaskii 114 NVNGEEIFIC GT-YYHKKSK REEMLDKLKN FESEAKNYKK ----KILMLH
P. furiosus 132 KGVYKDLEIH GMKYMSS-AW FEANKEILKR LF-RPT---- -DN-AILMLH
Consensus Q n h s - - - a h t p t d y l p e q f - - l p d f y d y v i w G H e H e c r i d p s y n p - - - f -
Conservation
RMSD: backbone
H. sapiens 218 QNRS---KHG STNFIPEQF- -LDDFIDLVI WGHEHECKIA PTKNEQQLFY
S. cerevisiae 214 QNHT---GHT NTAFLPEQF- -LPDFLDMVI WGHEHECIPN LVHNPIKNFD
S. pombe 223 QNHS---AHT PTSYLPESF- -IQDFYDFVL WGHEHECLID GSYNPTQKFT
C. thermophilum 214 QNHY---AHT PTGYLSENM- -LPDFLDLVI WGHEHECLID PKKNPETGFH
C. thermophilum (RBD) 436 ---------- ---------- ---------- ---------- ----------
M. jannaskii 159 QGIN---PYI PLDY-ELEHF DLPK-FSYYA LGHIHK-RIL ERFND---GI
P. furiosus 174 QGVREVSEAR GEDYFEIGLG DLPEGYLYYA LGHIHK-RYE TSYSG---SP
Consensus v v q p G S s i a t s l s p g e a v - - - - - - - - - - - - - - p K h f y i l d i - - g - - - -
Conservation
RMSD: backbone
H. sapiens 263 ISQPGSSVVT SLSPGEAV-- ---------- ----KKHVGL LRIK-GR---
S. cerevisiae 259 VLQPGSSVAT SLCEAEAQ-- ---------- ----PKYVFI LDIKYGE---
S. pombe 268 VVQPGSTIAT SLSPGETA-- ---------- ----PKHCGI LNIT-GK---
C. thermophilum 259 VMQPGSSIAT SLVPGEAV-- ---------- ----PKHIAI LSIT-GK---
C. thermophilum (RBD) 436 ---------- ---------- ---------- ---------- ----------
M. jannaskii 200 LAYSGSTEII YRNEYEDY-- ---------- -KKEGKGFYL VDF-SGNDLD
P. furiosus 220 VVYPGSLE-R WDFGDYEVRY EWDGIKFKER Y-GVNKGFYI VE--------
Consensus d f k m e k i p l - t v R p F i - - d i - l a - - - - - - - - - - - - - - k q - a - - - - l - e
Conservation
RMSD: backbone
H. sapiens 293 KMNMHKIPLH TVRQFFMEDI VLA---NHPD IFNPDNPKVT Q-AIQSFCLE
S. cerevisiae 290 APKMTPIPLE TIRTFKMKSI SLQ---DVPH LR----PHDK D-ATSKYLIE
S. pombe 298 DFHLEKIRLR TVRPFIMKDI ILS---EVSS I-PP-MVENK KEVL-TYLIS
C. thermophilum 289 SFEVEKIPLR TVRPFVIREI TLATDKRFKG LEKK--QDNR QEVTKR-LMQ
C. thermophilum (RBD) 436 ---------- ---------- ---------- ---------- ----------
M. jannaskii 236 ISDIEKIDI- ECREFV--EV NIK------- -------D-K KSFNE-AV-N
P. furiosus 260 DFKPRFVEIK -VRPFI--DV KIKG------ --------SE EEI-RKAI-K
Consensus k v e e m i - - a - - - - - - - - - - - - - - - - - - e - - p p k p l i r l r v d y s - - - - - - -
Conservation
RMSD: backbone
H. sapiens 339 KIEEMLENAE RER---LG-- -------NSHQ PEKPLVRLRV DYSG------
S. cerevisiae 332 QVEEMIRDAN EETKQKLADD GEGDMVAE-- LPKPLIRLRV DYSAPSNTQS
S. pombe 342 KVEEAITEAN AQWYEAQGTV PVV--ENEK- PPLPLIRLRV DYTG------
C. thermophilum 336 IVEEMIAEAN EMWRSLHEDS QDD--EDEEQ -PLPLIRLKV EYSS------
C. thermophilum (RBD) 436 ---------- ---------- ---------- ---------- ----------
M. jannaskii 266 KIER------ ---------- -------CKN KPVVFGKIKR E---------
P. furiosus 291 RLIPLI---- ---------- ---------- PKNAYVRLNI GWRK------

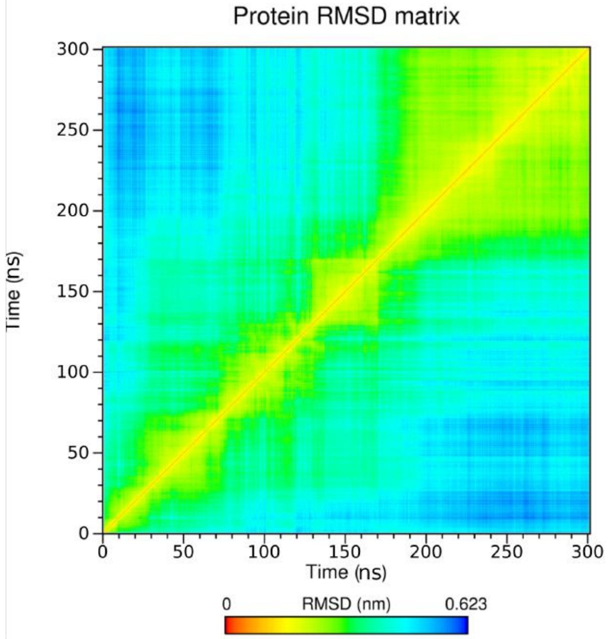

Consensus
Conservation
RMSD: backbone
H. sapiens
S. cerevisiae
S. pombe
C. thermophilum
C. thermophilum (RBD)
M. jannaskii
P. furiosus
----------- ---rFsnrf- vgrvan---- --dvv----q fyrkk-----
372 --G--FEPF- SVLRFSQKF- VDRVAN---- PKDII----H FFRHREQKEK
380 PID--YQ-VE NPRRFSNRF- VGRVAN---- GNNVV----Q FYKKRSPVTR
383 --G--YQ-TE NPQRFSNRF- VGRVAN---- ATDVV----Q FYLKK-----
377 PEGTKFE-VE NPQRFSNRF- AGKVAN---- QNDVV----H FYRKKTGTTR
436 ---------- ---------- ---------- ---------- ----------
284 ---------- ----FKP-WF -DTLKDKILI NKAIIVDDEF IDMPDN-VDI
311 P--------- ----FDLTE- IKELL----- --NV----EY LKIDTWRI--
---------- ---------- ---------- --l------- ----------
408 -TGEE----- ---------- ---------- ---------- ----------
418 -SKKSGINGT SISDRDVEKL FSESGGELEV QTLVNDLLNK MQLSLLPEVG
413 ---------- ---------- ---------- ---------- ----------
417 KPKEGKRELP EGIAEALE-- ---DSDSISV DALVQEFFAQ QSLKILPQAP
436 ---------- ---------- ----SDSISV DALVQEFFAQ QSLKILPQAP
317 ---------- ----ESLN-- ---------I KELLVDYANR QGID--GDLV
333 ---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- g--------- ----------
411 ---------- ---------- ---------- ---------- ----------
467 LNEAVKKFVD KDEKTALKEF ISHEISNEV- GILSTNEEFL RTDDAEEMKA
413 ---------- ---------- ---------- ---------- ----------
462 FGDAVNQFVS KDDKHAVEMF VMDSLSSQVR GLLQLDDDKI NEGLDSHIED
462 FGDAVNQFVS KDDKHAVEMF VMDSLSSQVR GLLQLDDDKI NEGLDSHIED
340 LS-LYKALLN NEN---WKEL LDEYYNTKFR G--------- ----------
333 ---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- --------
411 ---------- ---------- ---------- --------
516 LIKQVKRANS V-----RPTP PKENDETNFA FNX-----
413 ---------- ---------- ---------- --------
512 FRKVMEK-NF LSGQQKQAQR RRRFKEKAAA LEHHHHHH
512 FRKVMEK-NF LSGQQKQAQR RRRFKEKHHH HHH-----
366 ---------- ---------- ---------- --------
333 ---------- ---------- ---------- --------

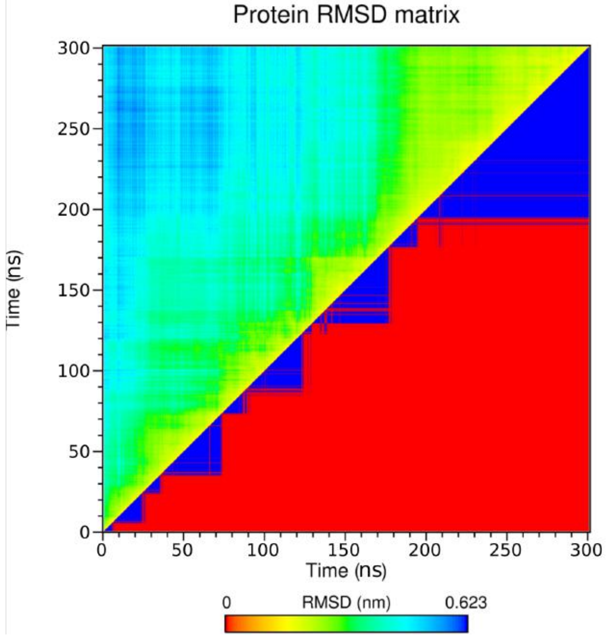
Appendix B

The initial model of the tetrameric structure of Mre11-Rad50 complex obtained by homology modelling is refined by molecular dynamics of 300 ns in NPT ensamble at temperature of 300 K and pressure of 1 atm. Simulation is performed using the AMBER99 force field and SPC/E water model. Principal component analysis (PCA) combined with cluster analysis are used to select the most representative structure during the simulation. FEL (contour plot) and percentage of existence (3D-histogram) are calculated using as reaction coordinates the projection of trajectory along PC1 vs. PC2. Clustering results of Gromos algorithm (see Chapter 20.9) are reported on the top of FEL, each structure is coloured according to the cluster to whom it belongs.

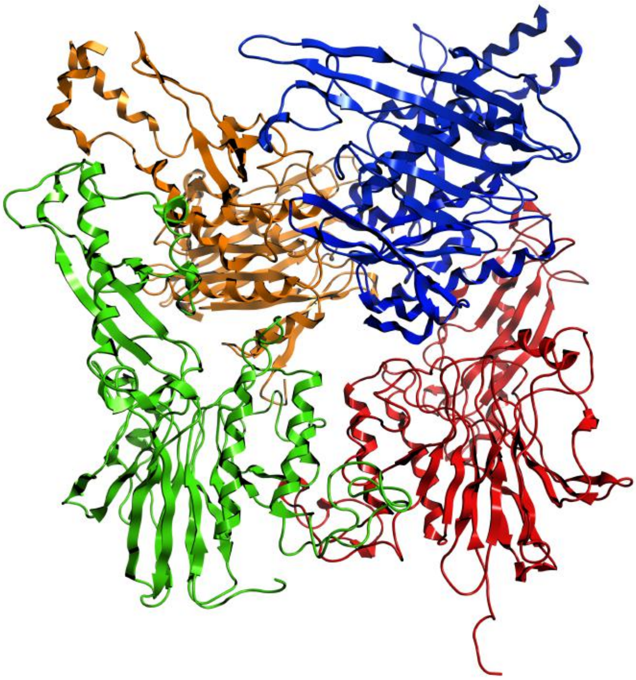


RMSD matrix of the trajectory of Rad50-Mre11 tetramer and clustering analysis performed on the RMSD matrix

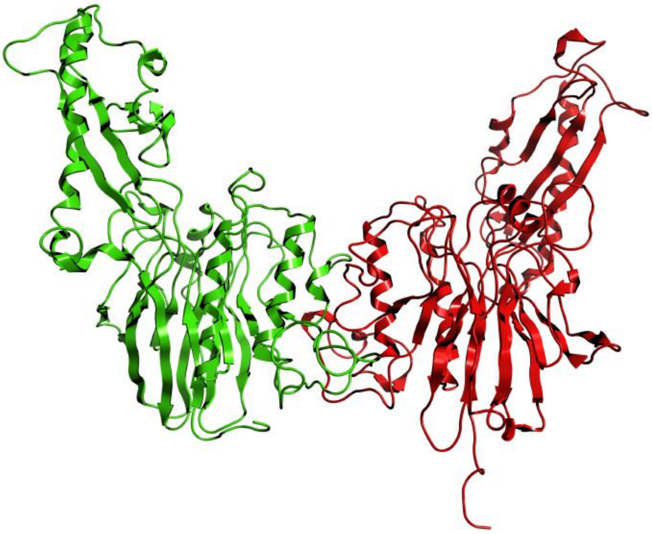




Structure taken from cluster-1 is selected as the most likely structures for the tetramer complex. In the following picture Mre11 subunits are reported in green and red, Rad50 subunits are represented in orange and blue.

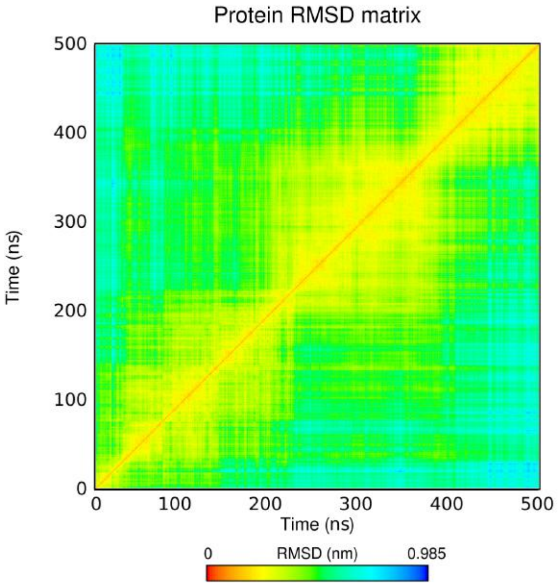


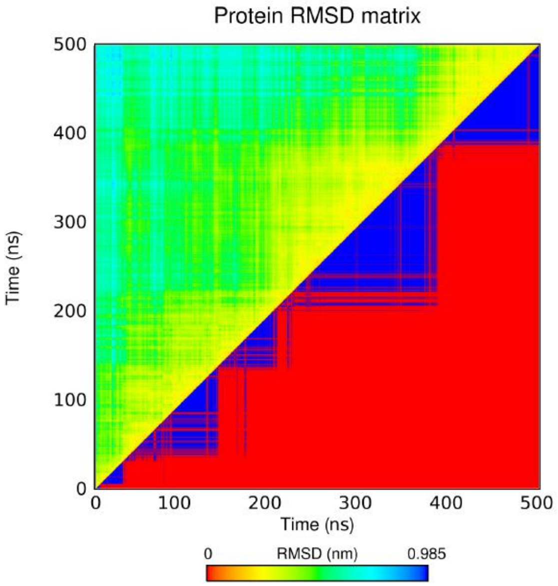
Mre11 dimer structure is extracted from the Mre11-Rad50 tetramer removing the Rad50 subunits.



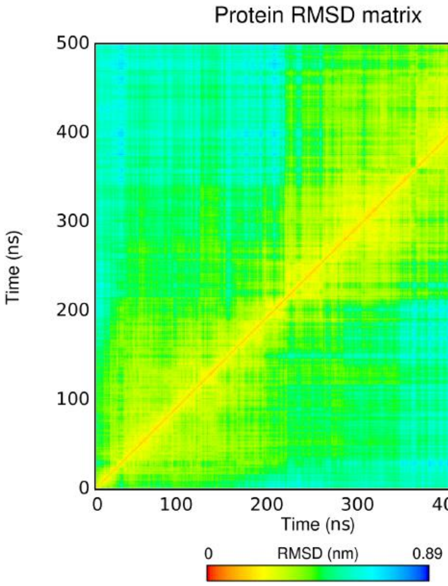
Appendix C

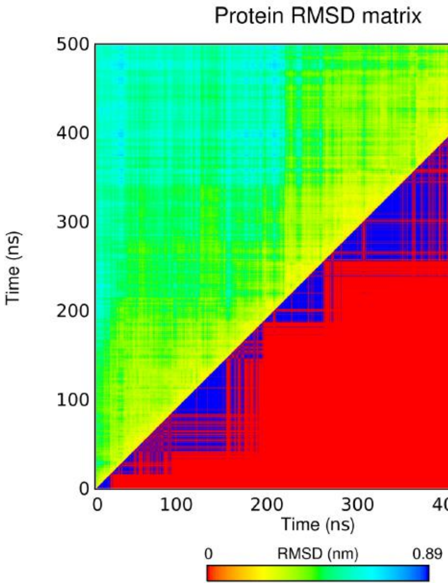
RMSD matrix of the trajectory of Mre11 wild type dimer and clustering analysis performed on the RMSD matrix





RMSD matrix of the trajectory of Mre11 mutant dimer and clustering analysis performed on the RMSD matrix





List of publications

The thesis is based on the following publications:

- **Jacopo Vertemara**, Claudio Greco, Luca De Gioia, Giuseppe Zampella. Quantum chemical study of the mechanism of the Mo-nitrogenase: characterization of structural and electronic properties of the activate state E_4. In preparation.
- Stephen Keable, **Jacopo Vertemara**, Oleg Zadvornyy, Brian Eilers, Karamatullah Danyal, Andrew Rasmussen, Luca De Gioia, Giuseppe Zampella, Lance Seefeldt, John Peters. Structural Characterization of the Nitrogenase MoFe Protein with the Substrate Acetylene Trapped Near the Active Site. *Journal of Inorganic Biochemistry*, doi: 10.1016/j.jinorgbio.
- **Jacopo Vertemara**, Melissa Zastrow, Vincent L. Pecoraro, Luca De Gioia, Giuseppe Zampella. Mechanistic investigation on ester hydrolysis catalyzed by de novo designed peptide TRIL9CL23H and comparison to carbonic anhydrase. In preparation.
- Elisa Gobbini, Corinne Cassani, **Jacopo Vertemara**, Weibin Wang, Fabiana Mambretti, Patrick Sung, Renata Tisi, Giuseppe Zampella, Maria Pia Longhese. The MRX complex enhances Exo1 resection activity by altering DNA-end structure. *The EMBO Journal*, Submitted.

Additional publications not discussed in this thesis:

- Luca Bertini, Marta Alberto, Federica Arrigoni, **Jacopo Vertemara**, Piercarlo Fantucci, Maurizio Bruschi, Giuseppe Zampella, Luca De Gioia. On the photochemistry of Fe2(edt)(CO)4(PMe3)2, a [FeFe]-hydrogenase model: A DFT/TDDFT investigation. *International Journal of Quantum Chemistry.* doi 10.1002/qua.25537

- Cristina Visentin, Francesca Pellistri, Antonino Natalello, **Jacopo Vertemara**, Marcella Bonanomi, Elena Gatta, Amanda Penco, Annalisa Relini, Luca De Gioia, Cristina Airoldi, Maria Elena Regonesi, Paolo Tortora. Epigallocatechin-3-gallate and related phenol compounds redirect the amyloidogenic aggregation pathway of ataxin-3 towards non-toxic aggregates and prevent toxicity in neural cells and Caenorhabditis elegans animal model. *Hum Mol Genet.* 2017 Sep 1;26(17):3271-3284. doi: 10.1093/hmg/ddx211.

- Corinne Cassani, Elisa Gobbini, **Jacopo Vertemara**, Weibin Wang, Antonio Marsella, Patrick Sung, Renata Tisi, Giuseppe Zampella, Maria Pia Longhese. Structurally distinct Mre11 domains mediate MRX functions in resection, end-tethering and DNA damage resistance. *Nucleic Acids Research*, Submitted.

- Diego Forni, Rachele Cagliani, Chiara Pontremoli, Uberto Pozzoli, **Jacopo Vertemara**, Luca De Gioia, Mario Clerici, Manuela Sironi. Evolutionary analysis provides insight into the origin and adaptation of HCV. *Journal of infectious diseases*, Submitted.