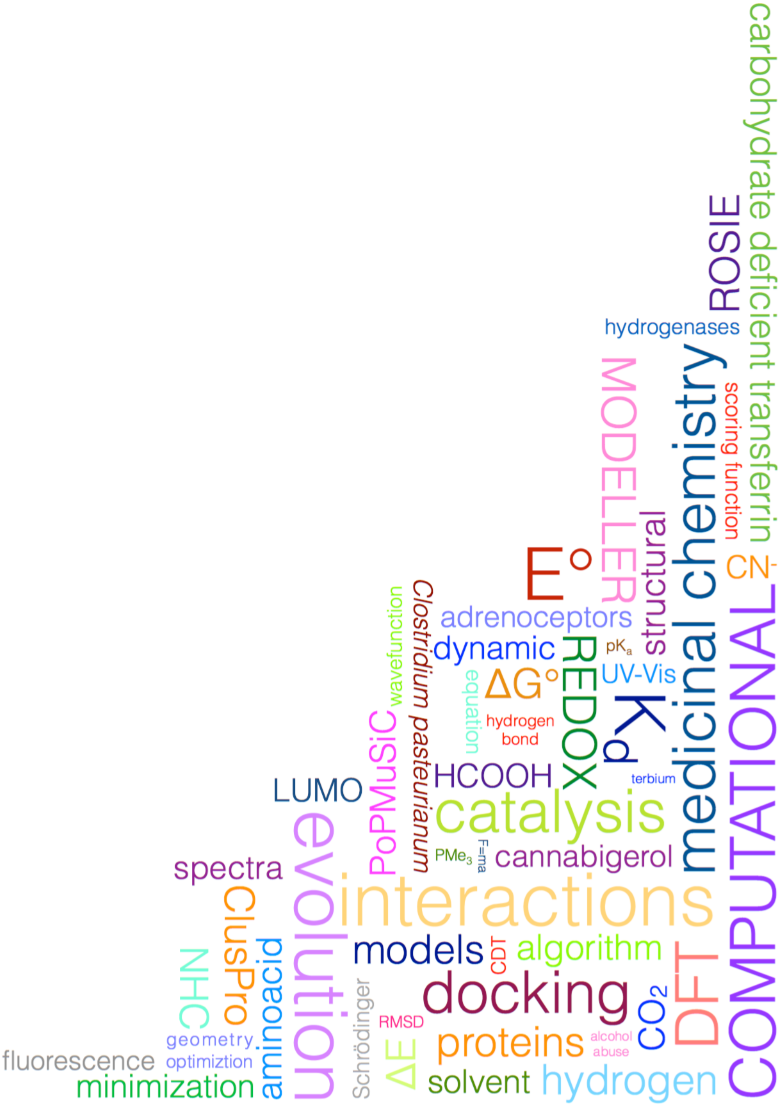
Universita' degli Studi di Milano-Bicocca
Dipartimento di Biotecnologie e Bioscienze
Dottorato di ricerca in Biotecnologie Industriali-XXVIII ciclo

Computational approaches for the study of biotechnologically-relevant macromolecules.

Coordinatore: Prof. Marco Vanoni
Tutor: Prof. Luca De Gioia
Co-tutor: Dott. Giuseppe Zampella

Giulia Filippi
Matr. 711551

Anno accademico 2014/2015


carbohydrate deficient transferrin
ROSIE
hydrogenases
MODELLER
medicinal chemistry
scoring function
CN-
E°
structural
adrenoceptors
dynamic
pK$_a$
UV-Vis
REDOX
ΔG°
K$_d$
Clostridium pasteurianum
wavefunction
equation
hydrogen bond
terbium
LUMO
PoPMuSiC
HCOOH
catalysis
COMPUTATIONAL
spectra
evolution
PMe$_3$
F=ma
cannabigerol
interactions
ClusPro
NHC
aminoacid
models
CDT
algorithm
DFT
Schrödinger
docking
CO$_2$
RMSD
geometry
optimiztion
ΔE
proteins
alcohol abuse
fluorescence
minimization
solvent
hydrogen

Summary

Computer-based techniques have become especially important in molecular biology, since they often represent the only viable way to understand some phenomena at the molecular level and gain insights into the behaviour of a biological system as a whole. The complexity of biological systems, which usually needs to be analyzed on different size- and time-scales and with different levels of accuracy, requires the application of different approaches. Computational methodologies applied to biotechnologies allow a molecular comprehension of biological systems at different levels of depth.

Ab-initio techniques based on quantum mechanics (QM) allow the study of enzymes and organometallic models at sub-atomic levels keeping into account electronic effects on stereochemistry and chemical reactivity. We set to study [FeFe]-hydrogenases, enzymes able to both produce and oxidize H_2 at high rate. The study was focused to better elucidate some redox states of the cofactor during catalysis. The principal aim of this work was to take advantage of hydrogenases biomimetic complexes to gain further inside on the catalysis of the enzyme, and pinpoint the structural and stereo-electronic features necessary to improve the efficiency of the synthetic models. H_2 is a desirable fuel but unfortunately, the usage of this gas is somehow problematic due to its physical properties, leading to safety concerns and low energy density. A possible way to overcome these problems is to store H_2 in safe and valued added chemicals. In this perspective we studied the catalytic mechanism of the first iron-containing synthetic complex able to catayze the chemical storage of H_2 and CO_2, converting it into HCOOH. QM methodologies were also used in a project in collaboration with the Dept. of Forensic Medicine at the University of Verona, aimed at the use of carbohydrate deficient transferrin (Tf) as marker of alcohol abuse. Tf is the protein deputed for the iron transport in the blood stream. Low glycosylated forms are known to be associated to alcohol abuse. Different spectroscopies were useful tools to discover the binding site of terbium and the best experimental conditions for terbium-Tf binding. To get more information about the active site, we optimized a method that allowed us to determine the molecular structure of the metal environment through QM computational techniques.

Docking techniques to study small-ligand protein binding are useful methods to predict the binding mode of a molecule to a receptor, in order to understand its mechanism of action and improve its activity. In addition, their combination with high throughput virtual screening (HTVS) allow to study a huge number of molecules in a relatively small amount of time, bypassing experimental costs. Here, we focused on pharmacological targets involved in different pathological mechanisms (nociception, cancer) to understand how the ligand is able to interact with the receptor and exert its pharmacological effect, and how to ameliorate its structure to increase the specificity.

Since the evolution of the human species exists a struggle for survival between host and pathogens, with measure and countermeasure to respectively infect and defend against infections. Positively selected sites on protein genes are the result of evolutionary pressure on certain aminoacidic residues that could be fundamental for host and pathogen infections. Protein-protein docking is a useful tool, together with computational stability analysis, to understand how residues variations modify the binding among different proteins in the immune system and how the proteins stability is affected.

In conclusion, the choice of the computational methods is what determines the level of the description of the molecular system. The study of biotechnologically relevant system with computational techniques is a powerful tool to gain insight into molecular properties that are otherwise not explorable by experimental techniques.

Contents

Chapter 1

From quantum to molecular mechanics: different approaches for different goals

> *We can still use the objectifying language of classical physics to make statements about observable facts. [...], but it is no longer possible to make predictions without reference to the observer or the means of observation.*
>
> *Niels Bohr*

As affirmed by Richard Feynman in 1964 in his Lectures on Physics, "*no subject or field is making more progress on so many fronts at the present moment, than biology, and if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms*". This assertion, beyond underline the highly dynamic nature of biological matter, stresses the remarkable importance of the study of atomic, molecular and macromolecular motions for the understanding of the fundamentals of life. Since the second half of the 20th century, besides classical experimental techniques, machine computations have played a critical role in science and engineering. Computer-based techniques have become especially important in molecular biology, since they often represent the only viable way to understand some phenomena at the molecular level and gain insights into the behavior of a biological system as a whole. The complexity of biological systems, which usually needs to be analyzed on different size- and time-scales and with different levels of accuracy, requires the application of different approaches.

Among the features we have to describe a system, its essential units or *particles* play a fundamental role in the choice of the computational technique. More in detail, the choice of *particles* and their number puts some limitations on what we are ultimately able to describe. If we choose atomic nuclei and electrons as our building blocks, we can describe atoms and molecules, but not the internal structure of the atomic nucleus. On the other hand, choosing the whole atom as building block, we can describe molecular structures, but not the details of the electron distribution. Reducing furthermore the resolution of the technique and opting for amino acids as the building blocks of our system, we may be able to describe the overall structure of a protein, but not the details of atomic movements.

The general aim of this dissertation is to apply and verify the usefulness and the limits of different computationals methods as tools to investigate the characteristics and properties of biotechnologically-relevant systems.

1.1 Quantum mechanical approach in bioinorganic chemistry

A considerable fraction of chemistry of life involves reactions that are catalyzed by metalloenzymes.[1] The scope of chemistry performed by these bioinorganic catalysts is very broad and spans through different types of reactions difficult to achieve in organic chemistry. Within the proteic scaffold, the metal ion, located in a pocket whose shape fits the substrate, usually can adopt several oxidation states and stabilize various reactive forms of substrates during the catalytic cycle, and hence provide a low energy path for chemical transformations. Very often, scientists try uncover the design principles of these natural devices, to study and develope fast and efficient biomimetic catalysts.

The term biomimetic chemistry was first used in 1972 to describe a branch of chemistry inspired by biological processes. Biomimetic chemistry covers a wide area of topics that include the synthesis and study of artificial enzymes and the study of biological precedents to direct the total synthesis of natural products. Generally this term refers to chemical catalysts that mimics certain key features of enzymes, including high enzyme substrate binding affinities, high turnovers and substantial rate accelerations relative to uncatalyzed reactions.

Since the reactions considered both in metallo-proteins and bioinspired catalysts are usually multi-step processes that involve short-lived and often highly reactive intermediates that frequently can decay along different scenarios, the study of these systems poses serious challenges to the scientific research.

Computational quantum chemistry is one of the various research techniques that can provide a detailed description of all species along the catalytic cycle in both enzymes and homogenous catalysts metal sites, including labile transition states and short-lived intermediates, which are frequently out of reach for temporary experimental techniques.[2]

In quantum chemistry, the fundamental rules and equations of quantum mechanics are applied to chemical issues. Quantum chemical models stem from the Schrödinger equation first brought to light in the late 1920's. It treats molecules as collections of nuclei and electrons, without any reference to their chemical bonds. The solution of the non-relativistic time-independent Schrödinger equation

$$E\psi = \hat{H}\psi \tag{1.1.1}$$

where $\hat{H}$ is the electronic Hamiltonian operator, E the energy of the molecule and ψ the wavefunction, which is a function of the position of the electrons and nuclei within the molecule, leads to the quantized energies of the chemical system and the associated wavefunctions. The wavefunction ψ can be approximated to the electronic state or, in other terms, to the configurations of the electrons in the molecular orbitals of a certain molecule. For a given wavefunction, the Hamiltonian operator can be calculated according to the following terms: 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. Unfortunately, the many-electrons Schrödinger equation can be solved exactly only for the simplest mono-electronic sistem: the hydrogen atom. Introduction of approximations is needed to provide a practical method that can be applied also to multi-atomic systems.

One way to simplify the Schrödinger equation for molecular systems is to assume that the nuclei do not move. In fact, the large difference in masses between electrons and nuclei allows for the decoupling of the instantaneous electron relaxation, with respect to the slow nuclear motion. This is called the *Born-Oppenheimer approximation*, and leads to an electronic Schrödinger equation where electron motion occurrs in the field of fixed nuclei. From all the terms listed above, the Hamiltonian lacks the nuclear kinetic energy and the nuclear-nuclear potential energy term is considered constant. Even after this expedient, electronic Schrödinger equation is still intractable and further approximations are required. As mentioned before, the Hamiltonian operator includes a term that is the sum of the repulsions between all electrons couples. Since repulsion energy is inversely proportional to electrons distance, it is necessary to be aware of the exact position of at least one of the two particles. This is hardly feasible in a multi-electronic system, and another simplification needed is the *molecular orbital approximation*, that entails the treatment of electrons as independent particles. Every electron is described as a wavefunction called spin-orbital

$$\phi_i(r_i, s_i) \tag{1.1.2}$$

that depends on spatial and spin coordinates, and is expressed as product of the spatial wavefunction, representing the molecular orbital, and the spin function (α or β). The new wavefunction (or Hartree product), given by the product of all the mono-electronic functions, expresses the probability of finding the electrons in a given space, with a certain spin. However, this latter is not completely satisfactory for electrons, because the wavefunction obtained from the product of all spin-orbital functions is not antisymmetric (the permutation of spatial and spin coordinates between two particles has to lead to the same wavefunction with different sign), as it must be from the Pauli exclusion principle. To overcome this problem, a linear combination of the both Hartree products of an anti-symmetric wavefunctions is introduced, giving back an expression, or *Slater determinant*, that describes the wavefunction of a multi-electronic system that satisfies anti-symmetry requirements and consequently the Pauli principle. In the *Hartree-Fock approximation* the wavefunction of a multi-electronic system is described by means of a single Slater determinant and corresponds to a single electronic configuration that derive from spin-orbitals at lower energy: the ground state. This approximation is suitable for describing a lot of molecular systems whose configuration is proximal to the equilibrium geometry. Depending on the restriction on spin-orbitals used, Hartree-Fock methodology can be classified as restricted or unrestricted. *Restricted Hartree-Fock* (or RHF) is applied to closed-shell systems, when all electrons are paired. Here it is imposed that α e β electrons occupy the same orbital, with different spin: there is a coincidence regarding the spatial volume where the electrons are located. On the contrary, *unrestricted Hartree-Fock* method (or UHF) is used for open-shell molecules where the numbers of electrons of each spin are not equal. In this case, different spatial molecular orbitals for the α and β electrons are applied.

The Hartree-Fock approximation leads to a set of equations, each taking into account the coordinates of a single electron. While they may be solved numerically, it is advantageous to introduce an additional approximation in order to transform them into a set of algebraic equations. It is reasonable to foresee that the one-electron solutions for many-electron molecules will closely resemble the one-electron solution for the hydrogen atom. Thus, the molecular orbitals are expressed as linear combinations of a finite set, a *basis set*, of prescribed functions known as basis functions. Generally these latter are atomic functions, or atomic orbitals, and every molecular orbital is expressed as a linear combination of a certain number of atomic functions. The choice of the appropriate basis set is extremely important in quantum chemical calculations, because the quality of all the results obtained will ultimately depend on the quality of the basis set used. The more basis functions are included in the molecular orbital calculation, the more accurate is the final representation. However, as might be expected, this results in an increase in computational time. The choice of the basis set is addressed by two different and necessary guidelines. First, they should reproduce the physics of the problem that ensures fast convergence as the number of basis function increases. Second, the practical way of selecting elegant and good functions should allow for rapid calculation of all integrals of interest to save computational time. The set of molecular orbitals leading to the lowest energy are obtained by a process referred to as a *self-consistent-field* or SCF procedure.

As mentioned before in the paragraph, Hartree-Fock models treat the motions individual electrons as independent of one another. To do this, they replace instantaneous interactions between individual electrons by interactions between a particular electron and the average field created by all the other electrons.

Because of this, electrons get in each others way to a greater extent than they should and there is an overestimation of the electron-electron repulsion energy and to too high a total energy. Electron correlation accounts for coupling and leads to a lessening of the electron-electron repulsion energy, and in the end to a lowering of the total energy.

One approach to the treatment of electron correlation is referred to as *density functional theory* or DFT. Density functional models have at their heart the electron density $\rho(r)$ as opposed to the many-electron wavefunction. Despite this, both similarities and differences can be found between wavefunction-based and electron-density-based methodologies. The essential building blocks of a multi-electron wavefunction are single-electron orbitals, that can be directly compared to the orbitals used in density functional methodologies. In addition, both the electron density and the many-electron wavefunction are constructed from an SCF approach.

The fundamentals of DFT derive from *Hohenberg-Kohn theorem*, that demonstrate how ground state associated electron density allow the calculation of a certain number of properties of the ground state itself. In detail, the ground state energy of the system E_0 is a *functional* (a rule that associates a number to a function) of the electron density

$$E_0 = E_0[\rho] \tag{1.1.3}$$

Unfortunately this theory expresses the existence and the uniqueness of the functional $E_0[\rho]$, but nothing is stated on its mathematical formulation: it is so necessary, even in this case, to perform some approximations.

Kohn-Sham theory states that the 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. Except for electrons kinetic energy, all components depend on the total electron density $\rho(r)$. Exchange-correlation energy is a functional of electron density that takes into account both exchange effects, the probability of finding two electrons with the same spin in the same point of space, and correlation effects, the probability of finding two electron with opposite spin. Electron density is expressed as summation of mono-electronic function named *Kohn-Sham orbitals* and since this term include the exchange-correlation functional it is necessary to define its mathematical form. Nowadays, differing functionals that use differing mathematical approximations to describe exchange-correlation energy have been introduced in everyday computational chemistry and they represent the attempt of make an educated guess on the real form of exchange–correlation energy.

The weakness of DFT is in the lack of a systematic procedure to progressively improve the functionals. However, a lot of work has been done following two different and collateral directions. First of all the functionals have to follow a set of formal properties and a behaviour physically acceptable. In addiction, the functional has to allow the reproduction of experimental data of an heterogenous set of molecular systems. This latter case has led in many instances to the introduction in the formalism of the functional of different arbitrary parameters, in order to achieve a better reproduction of a large set of experimental data. In addition to these theoretical weak points, a factor that can really determine the overall quality of a given quantum chemical calculation is the accuracy and the adequacy of the molecular model used. The choice of reliable and sufficiently large model may be thus very important for accurate description of enzymatic active sites. However the computational cost of correlated to *ab initio* methods mostly prevents their applications to very large systems.[3, 4, 5, 6, 7, 8, 9, 10, 11]

As extensively discussed in the previous paragraphs, with DFT is possible to analyze and study the electronic configuration of the ground state of a certain molecule. Time-dependent DFT, or TDDFT, extends the basic ideas of ground-state DFT to the treatment of excitations or more general time-dependent phenomena. TDDFT can be viewed an alternative formulation of time-dependent quantum mechanics but, in contrast to the normal approach that relies on wave-functions and on the many-body Schrödinger equation, its basic variable is the one-body electron density $\rho(r\ t)$. The standard way to obtain $\rho(r\ t)$ is with the help of a fictitious system of non-interacting electrons. The final equations are simple to deal with numerically, and are routinely solved for systems with a large number of atoms. These electrons detect an effective potential, the *time-dependent Kohn-Sham potential*. The exact form of this potential is unknown, and has therefore to be approximated. Two different situations can however be observed. If the time-dependent potential is weak, it is sufficient to draw upon linear-response theory to study the system (induced change in the density can be described as linearly dependent on the applied perturbation). In this way it is possible to calculate optical absorption spectra and circular dichroism spectra. On the other hand, if the time-dependent potential is strong, a full solution of the Kohn-Sham equations is required. A classical example of this regime is the treatment of atoms or molecules in strong laser fields. In this case, TDDFT is able to describe nonlinear phenomena like high-harmonic generation or multi-photon ionization.[12, 13, 10]

1.2 Understanding molecular interactions through molecular mechanics

The exponential growth of three-dimensional structures of biomolecules we are witnessing in these latest 20 years has completely revolutionized the approach of modern biology. A lot of structures available nowadays are proteins that play a prominent role in cellular biochemistry and physiology, and offer huge opportunities in the field of structure biology.

As told in the previous section, a complete description of the energy of a molecule can be achieved only with the solution of the Schrödinger equation. Altough this equation can be generally adopted to calculate the energy of every molecular system, its practical application in macromolecular structural biology field is negligible due to the intrinsic complexity of the *ab initio* computational methodologies. Molecular mechanics (or MM) instead, uses classical mechanics to model molecular systems and evaluate energy associated to different conformations.

In MM models the building blocks are atoms, meaning that electrons are not considered as individual particles. Thus, bonding information must be provided explicitly, rather than being the result of solving the electronic Schrödinger equation. In addition to bypassing the solution of the electronic Schrödinger equation, the quantum aspects of the nuclear motion are also neglected. The physical consideration of these structures in the same way as macroscopic bodies, allow their study through the application of Newton equation

$$F_i(t) = m_i \, a_i(t) \tag{1.2.1}$$

where F is the force acting on i-th atom at t, m is the mass and a is the acceleration. For time-independent phenomena, the problem reduces to calculating the energy at a given geometry. Often the interest is in finding geometries of stable molecules and/or different conformations, and possibly also interconversion between conformations. Molecules are described by a ball (atoms) and spring (bonds) model in MM methods, with atoms having different sizes and softness and bonds having different lengths and stiffness. MM found its complete expression in the *force field*, a set of funtions and parameters used to calculate the potential energy of a molecular systems in a given conformation as a sum of individual energy terms. Intuitively it is reasonable to expect that the functional form of potential energy accounts for a sum of terms, each describing the energy required for distorting a molecule in a specific fashion. The total potential energy U of the system can be written as

$$U = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross} \tag{1.2.2}$$

where E_{str} represents the energy function for stretching a bond between two atoms, E_{bend} represents the energy required for bending an angle, E_{tors} is the torsional energy for rotation around a bond, E_{vdw} and E_{el} describe the non-bonded interactions between atoms, and E_{cross} describes the coupling between the first three terms. Given such an energy function of the nuclear coordinates, geometries and relative energies can be calculated by optimization. Both E_{str} and E_{bend} represent a practical application of Hook's law, that explain the behaviour of elastic bodies in presence of a certain external force. While E_{str} takes into account the energy associated to vibrations around bond equilibrium lenght, E_{bend} count the energy associated to little variations around angle equilibrium. It has to be underlined that the term E_{str} loses significance when bonds lenghten to dissociation values. The E_{tors} term typically has multiple minima and thus cannot be modeled as harmonic oscillators. It describes part of the energy change associated with rotation around a B—C bond in a four-atom sequence A—B—C—D, where A—B, B—C and C—D are bonded. E_{tors} has contributions from both the non-bonded (van der Waals and electrostatic) terms and the torsional energy function must be periodic, since the bond is rotated 360° the energy should return to the same value. E_{vdw} is the van der Waals energy describing the repulsion or attraction between all couples of atoms that are not directly bonded. This term is zero at large inter-atomic distances and becomes very repulsive for short distances. A popular function that obeys these general requirements is the Lennard-Jones potential. On the other hand, E_{el} contribution is described by the Coulomb potential, and de-

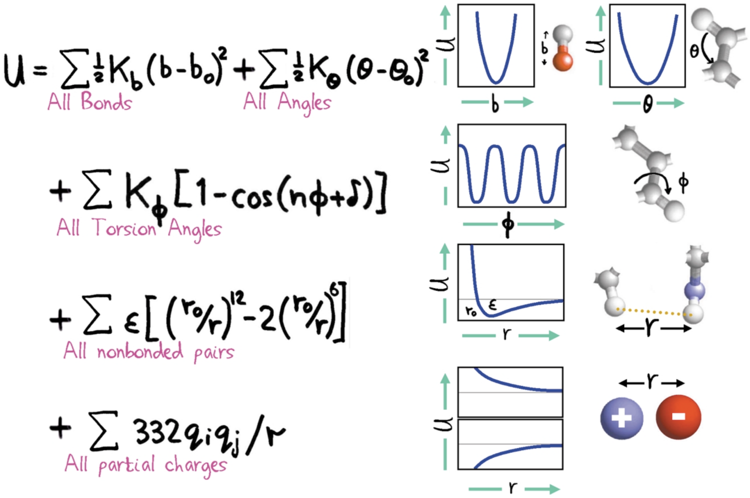


Figure 1.2.1: Total potential energy U of any molecule is the sum of simple allowing for bond stretching, bond angle bending, bond twisting, van der Waals interactions and electrostatics.

pends on the partial charges of the two atoms considered, on their distance and on the dielectric of the medium. Together E_{vdw} and E_{el} describes the non-bonded energy of the molecular system. The last term, E_{cross}, covers coupling between the five fundamental terms taken into account for the calculaton of the total potential energy of the system. Taken all together, the six terms that contribute to the determination of can be sum up as depicted in Figure 1.2.1.

The application of the principle of molecular mechanics founds one of the most intresting application in molecular *docking*, a methodology arisen in the latest years of the 20th century. The problem of docking is determining the best alignment of two molecules with respect to each other, typically trying to fit a small molecule into a large protein structure, but this also works between two different proteins. Considering the case of a small ligand with a protein structure, docking methodology intends to identify the geometrical transformations that has to be applied to ligand coordinates in order to minimize the potential energy of the ligand-protein complex. It is possible to imagine that if we consider the ligand completely flexible (rotable bonds has no torsional angles), the exponential number of possible conformations that can be adopted, make the research of a stable ligand-protein interaction extremely intractable. To overcome this obstacle, any docking program has to implement an efficient research algorithm that allow the user to explore all the conformations possible, and a function able to quantify the interaction energy between ligand and protein, and rank al the possible solutions. This function, called *scoring function*, very often is molecular mechanics force field based, and allow to estimate the energy of a ligand conformation. Here in docking methodology, forcefield are modified in order to take into consideration also entropic and desolvation effects. Indeed after the binding, the ligand and to a lesser extent the protein, undergo to a loss of entropy, energetically disadvantegeous, in favor of a partial or total loss of interaction with the solvent molecules. These aspects are of great importance in docking procedure, but can be neglected in canonical force field.[14, 15, 10, 16, 17]

1.3 PES topology and stationary points

Both the solution of the Schrödinger equation and a force field, represent an empirical approximation of the potential *energy surface* (or PES) of an atomic system. Indeed, having an equation or a function that describes the potential energy of a system starting from its atomic coordinates, it could be possible to draw a 3N dimensional surface (with N is equal to the number of atoms of the system) where every point it's defined by the potential enery of the system with a certain atomic configuration. This surface has, for every biological molecule, infinite dimensions. Nevertheless this representation finds its utility in the description of some molecular properties. There are some points on the PES that are particularly significant in energetic terms and they are called stationary points. Among these, we can distinguish *maximum*, *minimum* and saddle points. *Maximum points* represent unfavorable conformations of the molecule, while *saddle points* correspond to labile geometries, distinctive of the interconversion process between conformers, and also called *transition states*. The most energetically stable conformation on the PES is the *global minimum*. It can happen that a molecule acquires a conformation adequately stable and falls in a hole on the PES, surrounded by unsurmountable energetic peaks: these points are called *local minima*. The analogy generally used to describe the PES is that of a natural landscape, made up of its hills and valleys. Like gravity, the force field generates the force that drives and object plunged in the field to reach en equilibrium, where the sum of all the forces is null. Both quantum and molecular mechanics is interested in the study of these energetically stable points on PES. The process that take advantage of the PES to reach the energetic *minimum* starting from a certain conformation of a molecule is called *minimization*. This procedure is based on the following mechanism: given a set of atomic coordinates x_k, which define the position and conformation of a molecular system in cartesian space, and a function f that describes potential energy associated to every conformation k, coordinates values at which f displays a minimum are evaluated. This condition is fulfilled when

$$\frac{\partial f}{\partial x_k} = 0 \qquad \frac{\partial^2 f}{\partial x_k^2} > 0 \tag{1.3.1}$$

Since $\mathbf{x_k}$ is a vector composed by 3N variables x_k, f first derivative (also called *gradient*) is a vector f where every component is the partial derivative of f with respect to x_k.

One of the most used approach in energy minimization is the *steepest descent method*, where the direction in which the geometry is first minimized is opposite to the direction in which the gradient is largest (or steepest) at the initial point. Once a minimum in the first direction is reached, a second minimization is carried out starting from that point and moving in the steepest remaining direction. This process continues until a minimum has been reached in all directions to within a sufficient tolerance. In the *conjugate gradient method*, the first portion of the search takes place opposite the direction of the largest gradient, just as in the steepest descent method. However, to avoid some of the oscillating back and forth that often plagues the steepest descent method as it moves toward the minimum, the conjugate gradient method applies the correction of the gradients, evaluating the precedent direction. This allows the method to move rapidly to the minimum. The *Newton-Raphson method* is the most computationally expensive of all the methods here mentioned. It is based on Taylor series expansion of the potential energy surface at the current geometry. The correction of the geometry depends on both the first derivative (also called the slope or gradient) of the potential energy surface at the current geometry and also on the second derivative (otherwise known as the curvature). Although the computational expense, the Newton-Raphson method usually requires the fewest steps to reach the minimum.[18, 10, 16, 19]

Chapter 2

Lessons from Fe-containing metalloenzymes: hydrogen production and storage.

> *It's is not enough that you should understand about applied science in order that your work may increase man's blessings. Concern for the man himself and his fate must always form the chief interest of all technical endeavors [...]. Never forget this in the midst of your diagrams and equations.*
>
> *Albert Einstein*

The widespread adoption of fossil fuels in these latest years has been responsible for the increasingly high concentration of carbon dioxide (CO_2) in the atmosphere, which is directly connected to environmental concerns of primary importance, such as antropogenic global warming and oceans acidification. It has been postulated that since the industrial revolution, the concentration of CO_2 in the atmosphere has risen significantly from around 280 to more than 398 parts per million in 2015, and it is expected to continue to rise. [P. Tans, NOAA/ESRL, www.esrl.noaa.gov/gmd/ccgg/trends]

In a world concerned about the availability of fossils fuel reserves and CO_2 emissions, the research of clean and sustainable energy sources is becoming more and more essential.[20] In this scenario, production of molecular hydrogen (H_2) has become a target of primiry importance due to its nature as a clean, high-energy-density, and renewable energy carrier.[21] Nonetheless, current hydrogen production mostly involves energetically costly processes, such as gas reforming from hydrocarbons or water electrolysis.[22] To achieve a sustainable hydrogen economy, more efficient hydrogen production and storage technologies are required [22] and, in particular, the investigation of biological processes able to produce H_2 by clean low-carbon and low cost techniques is highly regarded. Unfortunately, nowadays the best synthetic catalysts for H_2 production and consumption utilize the expensive and relatively scarce noble metal platinum.[23] The hydrogenase enzyme hold out the promise of using cheap metals for this processes.

These enzymes were probably *invented* by nature during the earliest life of cells on our planet, when early-Earth atmosphere was H_2 rich, and are the means by which some microorganism take advantage of the utilization of this molecule as a source of reducing powers or to produce proton as final electron acceptors.[24] Hydrogenases catalyze the simple chemical reaction $2H^+ + 2e^- \rightleftarrows H_2$, or H_2 evolving reaction (HER). The reaction is reversible, and its direction depends on the redox potential of the components able to interact with the enzyme. In the presence of H_2 and an electron acceptor hydrogenases act as an H_2-uptake while, in the presence of an electron donor of low potential, it may use the protons from water as electron acceptors and release H_2.[24] In addition, depending on the location in the cell, hydrogenases may also be involved in establishing transmembrane proton gradients.[25] According to the metal composition of the active site, these proteins can be classified in [NiFe], [FeFe] and [Fe-only]-hydrogenases.[24, 26, 27] In the first part of this chapter I will focus on [FeFe]-hydrogenases, reporting briefly the main characteristics of this class of enzymes and the results obteined from different DFT studies on [FeFe]-hydrogenase bioinspired complexes. In the second part of this chapter, I will analize the results obtained from the analysis of an iron-catalyst for the reduction of carbon dioxide through hydrogenation to produce formic acid, a value added chemical useful also to safely stock H_2 and CO_2.

2.1 [FeFe]-hydrogenases for hydrogen production and oxidation.

Hydrogenases containing two iron atoms in their active site, or [FeFe]-hydrogenases, catalyze both H_2 oxidation and the reduction of protons to molecular hydrogen with significant efficiency. These enzymes achieve very high production rates, of up to 9000 molecules H_2 s^{-1},[28] and this makes them extremely interesting candidates concerning the development of applications based on biohydrogen or semi-artificial H_2 production systems.[29] [FeFe]-hydrogenases can be found in anaerobic eukaryotes, in anaerobic prokaryotes such as *Clostridia*, in anaerobic fungi and ciliates and in some green algae.[30, 25, 31, 32] In the last decade crystal structure analysis and X-ray absorption spectroscopy revealed the general cofactor architecture of [FeFe]-hydrogenases.[33, 34, 35, 36] The majority of [FeFe]-hydrogenases features a monomeric organization, contrasting with the basically dimeric structure concept of [NiFe]-hydrogenases.[37] The core domain of the protein is the 40 kDa structure that embeds the enzymatic active site, composed by a cubane [4Fe4S] cluster and a [$2Fe_H$] subunit. The smallest [FeFe]-hydrogenase ever identified is reduced to this catalytic core unit (structure type M1), and has only been found in *Chlorophycean* green algae.[32] Other [FeFe]-hydrogenase monomers show a high degree of modularity by featuring up to three additional N-terminal and up to two further C-terminal cluster binding motifs containing accessory FeS clusters that mediate electron transfer between the H-cluster and redox partners of the hydrogenase. Structural type M2 displays an additional bacterial ferrodoxin-like domain with two canonical cubane cluster. This subtype can be found in all major lineages of [FeFe]-hydrogenase phylogeny and is thus purposed to be the ancestral [FeFe]-hydrogenase group.[32] Subtype M3, typical for clostridial enzymes, comprises two further cluster binding motifs, one of these binding a [4Fe4S]-cluster ligated by three cysteinate groups and one histidyl residue, while the other one exhibits homology to plant type ferredoxins and contains a [2Fe2S]-cluster.[32] When H_2/H^+-turnover is coupled to NAD(P)H/NAD(P)-turnover, an additional C-terminal domain homologous to respiratory complex I can be found: this can be a ferrodoxin-like [2Fe2S] cluster in subtype M4 or a canonical [4Fe4S] cubane in subtype M5.[38]

As briefly introduced above, the active site, the so called H-cluster, consists of two coupled subclusters: a canonical [4Fe4S] cubane cluster linked through a cysteinyl sulphur group to the catalytic [2Fe2S] or [$2Fe_H$] subunit, where both H_2 production and oxidation takes place. The two Fe atoms, referred to as proximal [Fe_p] or distal [Fe_d] with respect to the cysteine residue are coordinated to unusual non-protein ligand molecules. These peculiar ligands, altough in the first analysis were identified as 4 CO in *Clostridium pasteurianum* CpI hydrogenase, are in fact two COs and two CN^-s. Another CO ligand can be

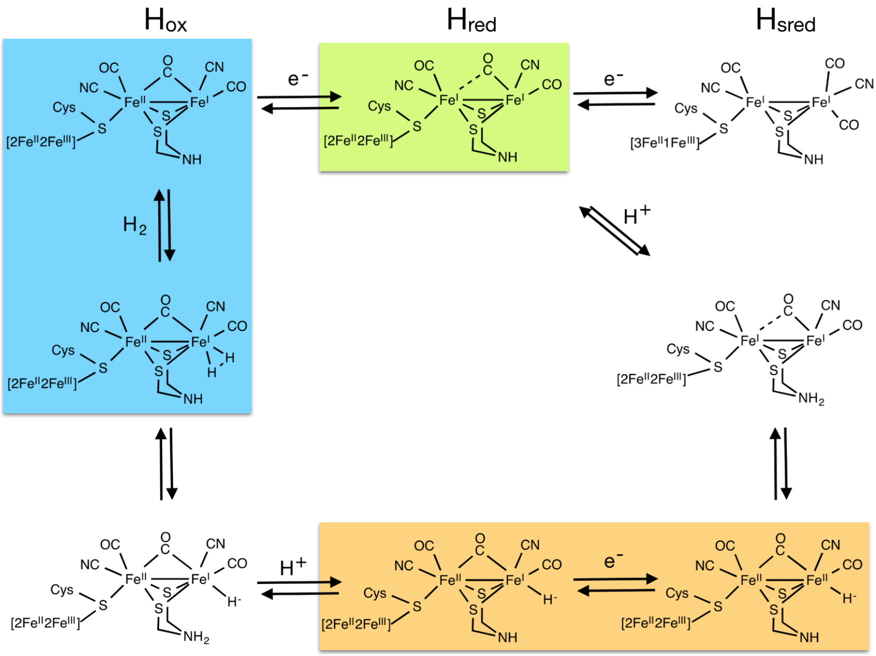


Figure 2.1.1: Catalytic cycle of [FeFe]-hydrogenases with key redox state indicated. The three coloured boxes indicate the catalytic steps studied in this work. Green: Section 2.1.1; orange: Section 2.1.2 and blue: Section 2.1.3.

found bridging the two iron atoms, or terminally bound to [Fe_d] depending on the redox state of the [$2Fe_H$] subunit.[33, 34, 35, 36] The inorganic sulfur atoms bridging the two irons are interconnected by a dithiolate bridge, whose central bridgehead atom has be identified as a nitrogen atom.[39] The amine group of the deriving azadithiolate (adt) could easily mediate, through a Walden inversion, the proton transfer from the proton transfer pathway to the open coordination site at [Fe_d].[39] The four cysteinates that coordinate the Fe sites of the [4Fe4S] subcluster, form the only covalent bonds between protein and H-cluster with one of the four cysteinates coupling the [$2Fe_H$] subunit and the cubane.[33] Besides these covalent interactions, non-covalent bonds between the proteic scaffold and the di-iron center stabilize and electrochemically tune the catalytic cycle (Figure 2.1.1).[40, 41, 21, 42, 43]

H_2 production and oxidation comprises two coupled electron and proton transfer steps during which the di-iron site passes through different oxidation states and ligand coordinations. Spectroscopically, two different redox state can be identified within the catalytic cycle: the paramagnetic H_{ox} state, exhibiting a mixed valence configuration at $[2Fe_H]$ site ($4Fe^{II}-Fe^{II}Fe^{I}$-[ ]), and the diamagnetic species H_{red} ($4Fe^{II}-Fe^{I}Fe^{I}$-[ ]).[40, 41, 21, 42, 43] Recently, a low potential redox state earlier regarded as inactive,[44] has been proposed as active and to be a further intermediate of the catalytic process of hydogenase: the super-reduced state or H_{sred} ($4Fe^{I}-Fe^{I}Fe^{I}$-[H^+]).[42, 45] At least for [FeFe]-hydrogenases of structure types M2 and M3, H_{sred} seems to be short-lived because the accessory [FeS]-clusters chain ease the release of the additional electron at the [$4Fe^{I}$]-cubane subcluster. As such accessory clusters are absent in M1 enzymes, here H_{sred} can be trapped *in vitro* as a stable state. According to the valence spectrum of H_{sred} , the binding site of the first proton prior to the second protonation step is still unclear, but an intermediate protonation of a nearby protein residue is suggested.[45] The active forms of the H-cluster are reversibly inhibited either by CO leading to the inactive oxidized state $H_{ox}CO$ or by formaldehyde.[30, 42, 46, 43]

As mentioned before, the H-cluster is embedded within the proteic scaffold, in a hydrophobic pocket with a rather complex fold of four helices and four loops.[33, 34] All cofactor-interacting amino acids are strictly conserved in all hydrogenases, with the exception of few cases in which the level of conservancy is anyhow high.[47] The conserved H-bond interaction between a lysine and the distal CN^- prevents the isomerization of the ligands at $[Fe_d]$.[35, 48, 42] The mutation of this residue to an asparagine led to a complete loss of activity of the enzyme due to the missing translocation of the $[2Fe_H]$ subcluster within the open binding site of the apoprotein, highlighting its role not only in catalysis but also in the maturation process.[49]

According to the [FeFe]-hydrogenases crystal structures available, the nitrogen atom of the adt linker is the terminal actor of the proton transfer chain between the solvent and the catalytic $[Fe_d]$, and it is thus pivotal for the catalytic mechanism.[42] This atom is able to interact with the nearest aminoacid (cysteine or methionine) through an hydrogen bond (N(H)−S or N−H(S)), modulating its basicity and its proton conductive ability. As expected, even a conservative mutation of this aminoacid abolishes or severely impairs the catalytic activity of the enzyme.[49]

Model studies of the [FeFe]-hydrogenase systems are necessary because the knowledge from X-ray absorption spectroscopies studies can't disclose all structural and redox properties of these biological machineries. Some features of these systems have been explored through synthetic chemistry and the different key areas of research on these systems, ranges from electrocatalysis and solid state devices, through hydride and mixed-valent systems, to total synthesis of H-cluster structures. Again unfortunately, some features remain ardueous to investigate due to the instability of some intermediates and the complexity of studying certain catalytic steps. DFT has been proved to be a useful tool to explore the catalytic features of both enzymes and homogenous catalysts metal sites. In addition computational studies allow to perform calculation on model systems to disclose some peculiar features, avoiding the costs and time employed in the process of synthesis.

This general overview on [FeFe]-hydrogenases will be followed in the forthcoming sections by more detailed introductions about the topics debated.

2.1.1 Biomimetic models of [FeFe]-hydrogenase H_{red} state that preserve the key structural features of the enzyme active site.

2.1.1.1 Introduction

Since the structural determination of the active site of [FeFe]-hydrogenases has been reported,[33, 34] the study of different derived complexes has driven a better understanding of the chemistry of these enzymes.[50, 51, 52, 53, 54] One of the still open issues concerns the structural features of the [$2Fe_H$] subcluster in the H_{red} redox state. The [Fe^IFe^I] complexes synthesized to date, fail to mimic the precise orientation of the diatomic ligands around the [2Fe2S] core that is observed in the reduced form of the enzyme.[39] [Fe^IFe^I] synthetic complexes are usually characterized by an eclipsed geometry of the ligands coordinated to [Fe_d], which has never been observed in the enzymatic catalytic site. Instead, the native [$2Fe_H$] subcluster is thought to be characterized by a rotated configuration of the ligands at [Fe_d], which exhibits an essentially square pyramidal coordination with a vacant coordination site in *trans* to the μ-CO group (see Figure 2.1.2).[39]

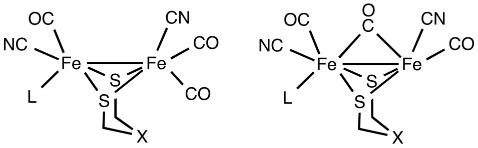


Figure 2.1.2: Schematic representation of eclipsed (left) and rotated (right) isomers.

This unique orientation of the diatomic ligands about the distal iron of the active site of [FeFe]-hydrogenases could promotes H^+ acceptance, and the progression of the catalytic cycle. Experimental and computational studies have suggested that the differences between the [$2Fe_H$] subcluster and synthetic models might be related and ascribed to enzymatic coordination sphere. The proteic scaffold of [FeFe]-hydrogenases is in fact able to create a unique network of electronic effects and hydrogen bonds with the di-iron cluster that is missing in the small synthetic analogues. DFT calculations have been carried out to evaluate how the structural features of the [$2Fe_H$] cluster derived complexes can be affected varying the first coordination sphere of ligands, suggesting that the combination of a sterically demanding dithiolate linker and asymmetric substitution of the CO ligands is particularly effective [35, 48] to obtain synthetic diiron complexes with structural properties similar to the [$2Fe_H$] active site.[55] More recently, a synergy of crystal packing effects and agostic interactions between the dithiolate bridge and an iron center were shown crucial to stabilize the rotated conformation in [Fe^IFe^I] biomimetic models.[56] One way to reproduce the key environmental factors present in the enzyme would imply the design of supramolecular complexes able to mimic the protein environment, where the position of hydrogen bond donors is similar to that observed in the active site. With the aim of evaluating whether the establishment of hydrogen bonding at the level of CN^-s is sufficient to favor rotation of ligands around one of the Fe centers, in analogy with experimentally observed geometry of the [$2Fe_H$] subcluster in H_{red}, the conformational properties and stabilities of a set of different isomers, which define a family of a basic biomimetic models of the di-iron subcluster have been studied. Then it has been investigated how a supramolecular amine-substituted porphyrin scaffold can affect their structural properties and relative stabilities.

2.1.1.2 Methods

All DFT calculations have been carried out gas-phase with the TURBOMOLE suite of programs,[57] applying the resolution of identity technique,[58] using the all-electron TZVP basis set [59] and the BP86 pure functional.[60, 61] This level of theory has been shown suitable and reliable for the investigation of [FeFe]-hydrogenase models. [62, 63, 48, 56, 64] In addition, correction for dispersion interaction was applied (DFT-D2).[65] Stationary points of the energy hypersurface have been located by means of energy gradient techniques, and full vibrational analyses have been carried out to further characterize each stationary point. In order to enclude entropy, zero-point energy (ZPE) and enthalpy contribution into the self-consisted field energy (E_{SCF}), free energy values (G) were calculated considering three different contributions to the total partition function ($q_{translational}$, $q_{rotational}$ and $q_{vibrational}$).[19] The T and P values were generally set at 298,15 K and 1 bar respectively. Because of the high values of the SCF wavenumbers arising from the harmonic approximation, a scaling factor of 0.9914 was applied.[57]

2.1.1.3 Results and discussion

The diiron dithiolates systems here considered to scan the potential energy surface is $[Fe_2(pdt)(CO)_3(CN)_2PMe_3]$ (pdt=propanedithiolate): while the pdt ligand was kept fixed between the two iron atoms, positions of COs, CN^-s and PMe_3 were exchanged in all possible manners to generate a set of 22 models (see Appendix). Among the 22 compounds, three different groups can be identified: **1-7** models feature an apical PMe_3 at $[Fe_p]$, while **8-14** and **15-22** models feature a basal PMe_3 at $[Fe_p]$ and $[Fe_d]$ respectively (Figure 2.1.3).

Model **1** (Figure 2.1.3) features the presence of one CN^- coordinated to each of the iron atoms, disposed in *trans* position with respect to the axis passing through Fe atoms (not differently from the case of the enzyme active site). In addition, the phosphine group is bound to the $[Fe_p]$ center in apical position, in full analogy with the disposition of the [4Fe4S]-subcluster in the H-cluster. Given such structural similarities with [FeFe]-hydrogenases catalytic site, and in consideration of the absence of any μ-CO isomer tightly reproducing the coordination geometry of $[2Fe_H]$, model **1** is taken as a reference point to discuss structure and relative stability of all the other isomers localized in our PES (see Table 2.1).

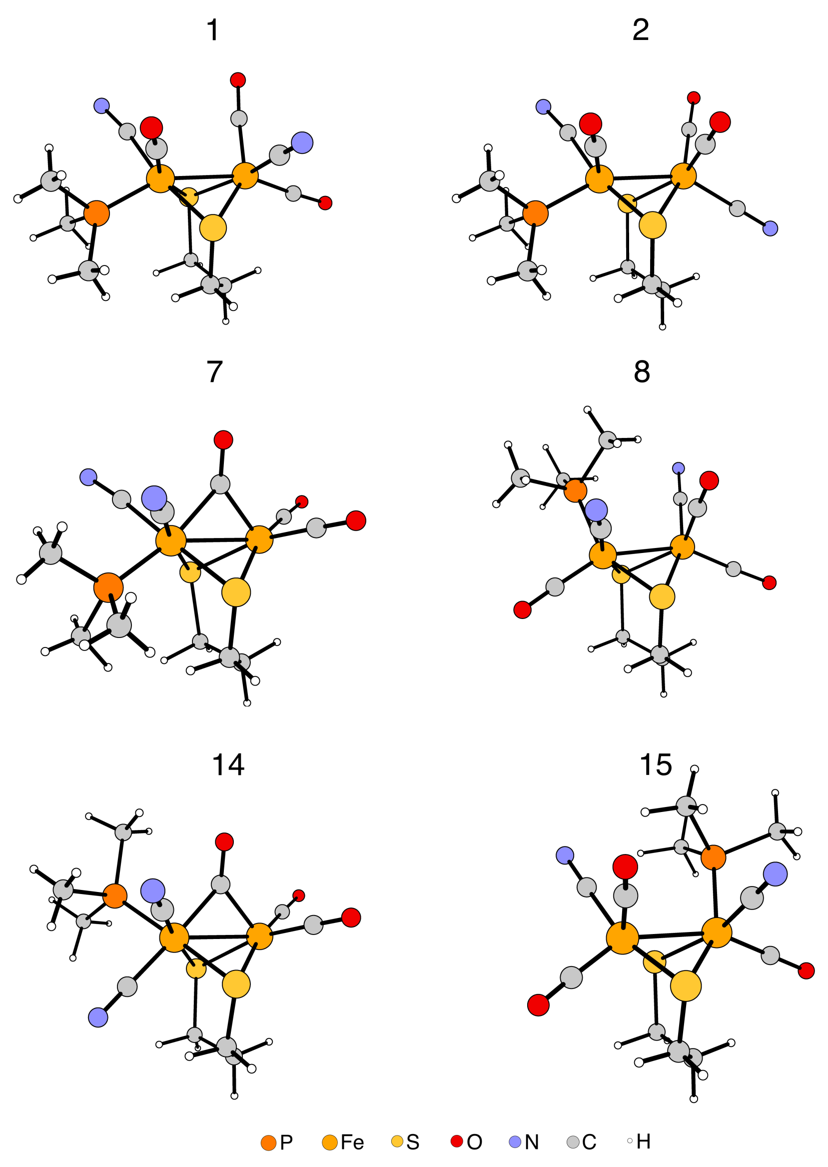


Figure 2.1.3: Optimized structures of selected diiron models.

	model	ΔG (kcal mol^{-1})	µ-CO
Apical PMe$_3$ at [Fe$_p$]	**1**	0.0	-
	2	-2.1	-
	3	-2.1	-
	4	+1.0	-
	5	+9.1	-
	6	+6.9	-
	7	+2.9	yes
Basal PMe$_3$ at [Fe$_p$]	**8**	+9.8	-
	9	-5.6	-
	10	-7.1	-
	11	-6.9	-
	12	-3.4	-
	13	-3.9	-
	14	-3.7	yes
Basal PMe$_3$ at [Fe$_d$]	**15**	-10.0	-
	16	-9.2	-
	17	-9.9	-
	18	-8.0	-
	19	-5.2	-
	20	-1.6	-
	21	-2.6	-
	22	-4.8	-

Table 2.1: Free energy differences of models **2-22**, taking the energy of **1** as reference point. The presence of a µ-CO between the two iron atom is here highlighted.

Within the first group of models, only **2** and **3** turned out to be lower in energy than the reference structure (ΔG = -2.1 kcal mol^{-1}), although the values obtained are close to the uncertainty of DFT method. Model **4** is isoenergetic with **1**, while all the others (**5**, **6**, **7**) were disfavoured (ΔG from +2.9 to +9.1 kcal mol^{-1}). Among them, optimization of model **7** brought to a µ-CO structure. This model displays an unsymmetrical configuration at the iron atoms, with all the σ-donors bound at the same Fe center. This configuration leads one of the π-acceptor groups to be rotated, bridging the two metal atoms and stabilizing the structure.[55] The group featuring the basal phosphines at [Fe$_p$] gather models more stable than **1** (ΔG from -3.4 to -9.8 kcal mol^{-1}). While models **8**, **9**, **10**, **11**, **12** and **13** show an all-terminal disposition of CO and CN$^-$ groups, model **14** (Figure 2.1.3) displays a bridging carbonyl group. Such picture is analogous with model **7**: again **14** showed all the σ-donors bound at the same Fe center, but turned out to be significantly more stable (ΔG = +2.9 kcal mol^{-1} *vs* ΔG = -3.7 kcal mol^{-1} respectively). In the third group of compounds exhibiting a basal phosphines at [Fe$_d$] center, all the structures turned out to be more stable or at least isoenergetic to model **1** (ΔG from -1.6 to 10.0 kcal mol^{-1}). In this group no µ-CO were identified. In this set of 22 compounds analyzed the only µ-CO models identified are the ones in which both cyanides are bound to the metal atom. No CO-bridged minimum can be found in models resembling the [2Fe$_H$] cluster, with one CN$^-$ at each Fe atom (Figure 2.1.3).

With the aim of evaluating whether the establishment of H-bonds at the level of CN$^-$s is sufficient to favor rotation around one of the Fe centers and induce the formation of a µ-CO, we studied a series of derived supramolecular assemblies obtained from the more stable models of Table 2.1 (**1-3** and **8-22**) in complex with an amine-substituted porphirin derivative (Figure 2.1.4). Results are shown in Table 2.2.

The presence of the porphyrin scaffold allows the diiron model to assume a µ-CO structure in only two minima: **A** and **B** (Figure 2.1.5). Model **A** in particular features an hydrogen-bond between a CN$^-$ ligand and one of the two amine group, as well as dispersive interactions between the phosphine ligand and the porphyrin ring. In this model the position of the carbonyl ligand between the two iron atoms is semi-bridging, with a [Fe$_p$]-C(CO) bond significatively longer than [Fe$_d$]-C(CO) distance. This ligand disposition on the di-iron subcluster resemble the H$_{red}$ form of the [2Fe$_H$] cluster.[39] A similar structure disposition

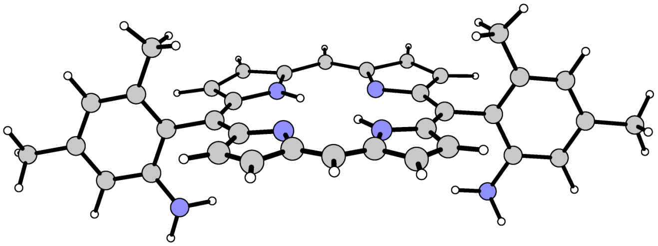


Figure 2.1.4: Structure of the amine substituted porphyrin derivative considered

model	ΔG (kcal mol^{-1})	μ-CO
A	0.0	yes
B	+12.1	yes
C	-0.4	-
D	+9.8	-
E	+1.5	-
F	+7.1	-
G	+9.2	-
H	+8.8	-
I	+11.8	-
L	+5.9	-
M	-1.3	-
N	+3.8	-
O	+4.4	-
P	+5.2	-
Q	+11.0	-
R	+4.9	-
S	+4.4	-
T	+1.3	-
U	+7.5	-

Table 2.2: Free energy differences of models **B-U** taking the energy of **A** as reference point. The presence of a μ-CO between the two iron atom is here highlighted.

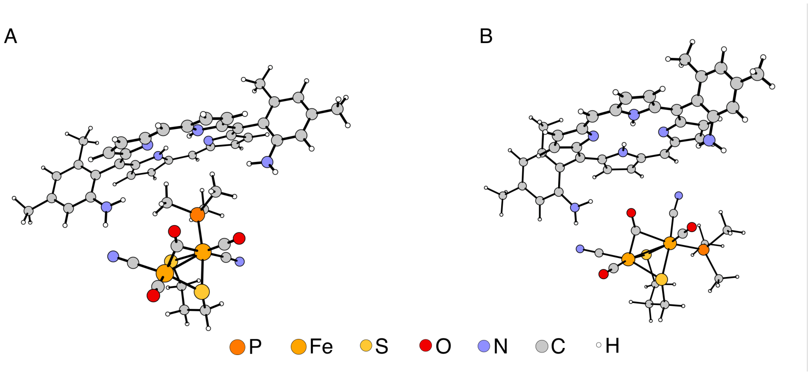
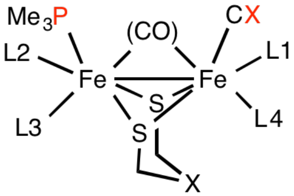


Figure 2.1.5: Optimized structures of selected porphirin-containing models featuring a μ-CO.

can be found also in **B**, although this model is +12.1 kcal mol^{-1} higher in energy than model **A**. In both **A** and **B** a vacant coordination site can be identified in *trans* position with respect to the bridging ligand, again in analogy with the reduced form of the H-cluster.[39] Among the other models, no one showed a μ-CO.

In order to better understand the reasons for which binding to the porphyrin favors a rotated μ-CO core, the steric requirements for the presence of a bridging ligand were investigated. The choice of such structure stems from the fact that polipeptide assemblies fail to mimic the relatively rigid tertiary structure of the folded protein. Analysis of the conformational space of the simple **1-22** models set presented in Table 2.1 and in Appendix shows that the distance between terminal ligands belonging to each of the two coordination spheres increases significantly when going from all-terminal to μ-CO adducts. The P(PMe$_3$)-O(CO) interatomic distance (see Figure 2.1.6 for a schematic representation) is larger in the μ-CO adducts such **14** (5.75 Å), while it is shorter (below 4.00 Å) in all other models featuring eclipsed conformation of ligands and a basal disposition of PMe$_3$.

The same results can be found taking into consideration P(PMe$_3$)-N(CN$^-$) distance in models **8** and **12** (3.92 Å in both cases). Taking all these informations, even though the reason why **14** assumes a bridging-CO conformation is esentially of electronic nature, the analysis illustrates that the elongation of interatomic distances between ligands belonging to the two coordination spheres represents a key steric requirement to accommodate a μ-CO group. A further



X = O in 9, 10, 11, 13, 14
X = N in 8, 12

Figure 2.1.6: Schematic representation of selected diiron models with highlighted P(PMe$_3$)-O(CO) distance.

support to this hypothesis is that also in **A** the distance between the phosphine group and the basal ligand in *cis* with respect to the Fe-Fe axis is relatively large (the P(PMe$_3$)-N(CN$^-$) distance is 5.40 Å).

2.1.1.4 Conclusions

To sum up, the scan of the isolated di-iron cluster evidenced that the bridging disposition of one of the CO ligand, a typical feature of the H$_{red}$ redox state, characterizes only models in which phosphine and CN$^-$ ligands are all bound to the same metal center. Minima with spatial orientation of ligands more closely related to the one observed in the enzyme active site could be located when the models were super-structured by the presence of a di-amine substituted porphyrin ring, able to behave as a scaffold that donates hydrogen bonds to the cyanide groups. Inspection of the structural features of such [2Fe2S]-porphyrin supramolecular models, indicate that the usage of rigid scaffolds covalently linked to cyanide-bound biomimetic diiron models can lead to the stabilization of μ-CO geometries resembling the one disclosed by crystallographic studies of H$_{red}$.

2.1.2 Dissection of the reduction step in H$_2$ catalytic production by [FeFe]-hydrogenase inspired models.

2.1.2.1 Introduction

The unique structure of the [FeFe]-hydrogenase active site encouraged the research of novel bioinspired complexes, in order to better figure out structure–activity relationships and ultimately design new synthetic catalysts. The reduction of protons, once these have bound the cofactor in its $[Fe^{I}Fe^{I}]$ redox state, is a key step in HER both in enzyme catalysis and i synthetic analogues. Here the focus is on the first reduction occurring during the catalytic production of H_2: $[Fe^{II}Fe^{II}]$-$[H^-]$$\rightleftharpoons$$[Fe^{II}Fe^{I}]$-$[H^-]$. The importance of mixed-valence species like $[Fe^{II}Fe^{I}]$-$[H^-]$ resides in their likely intermediacy in both the enzyme turnover and electrocatalysis by synthetic [FeFe]-hydrogenases analogues.[66, 67] As mentioned in the previous sections, the vacant site identified in the redox state H_{red} at $[Fe_d]$,[39] allows the proton binding in a terminal fashion (t-H).[68, 69, 70] Its corresponding bridging form (μ-H) however has been reported to be thermodynamically more stable.[48] Thus it is conceivable that a mechanism of rearrangement could, in principle, turn the t-H isomer into the corresponding μ-H form. Here, a key role in the [FeFe]-hydrogenases active site is played by the two aforementioned CN^--enzyme interactions, that hinder the isomerization at $[Fe_d]$ preventing the formation of a less reactive μ-hydride.[35, 48, 42] Despite the biological significance of the t-H species, their experimental detection in synthetic analogues has been rare, and requires special experimental condition because of their spontaneous tendency to isomerize.[71] An intimate mechanism able to describe the rearrangment of t-H isomer to its μ-H counterpart has been proposed in a DFT computational investigation illustrating the energetic viability of the process.[72] This reaction mechanism has been recently postulated to be absent in the enzyme cofactor in a recent work by Reiher's group.[73] The quantum mechanical study performed on a large model of the [FeFe]-hydrogenases active site, which took into account also the interactions of the H-cluster with the protein environment, illustrated that the formation of a thermodynamically more stable μ-H is kynetically hindered due to large energetic barriers. In addition, they explained that the surrounding amino acidic environment is crucial to prevent the μ-H formation, enabling a fast proton reduction with low barriers.[73] These data reinforce the aforementioned idea that too stable bridging-hydrides would form a thermodynamic sink also in the enzyme if a kinetic device was not present to freeze the t-H to μ-H isomerization. Diferrous species bearing t-H ligand were unknown since the first X-ray diffraction characterization performed by Rauchfuss' group.[74] The study of $[HFe_2(edt)(PMe_3)_4(CO)_2]$ [edt=ethanedithiolate] derivatives with hydride reagents at very low temperature provided a new approach to synthesize an active site model bearing a critically active intermediate. The t-H identified was indeed able to reacts with Bronsted acids to give H_2, in contrast to the bridgind counterpart.[74] Terminal coordination of hydrides was once again detected by Ezzaher et al. in solution at low temperature.[75] In these conditions, the isomerization process leading to bridging isomers is certainly slow to occur. In the same study, it was shown that protonation of the asymmetric complex $[Fe_2(pdt)(CO)_4(dppe)]$ [dppe=1,2-$C_2H_2(PPh_2)_2$] at two different temperatures (−50 and −75 °C) led to the formation of two different t-Hs, on the $(CO)_3$ Fe atom or on the (CO)(dppe) Fe atom, respectively.[75] Through different spectroscopic techniques, t-Hs were also characterized for a family of complexes with bulky bis-phosphine ligands [dppv=1,2-cis-$C_2H_2(PPh_2)_2$].[76, 77] Here, the lower isomerization to the μ-H congener is ascribable to the stericaly crowded Fe atoms and the hindrance of the dppv ligands. In one of these studies, the persistence of the terminal isomer was such to allow its redox potential measurement, along with that of μ-H. Analysis on pdt and adt derivatives of species investigated in the same study showed that t-Hs reduce at potentials of 100−200 mV less negative than their bridging counterparts ($\Delta E°_{tH-\mu H} > 0$).[77] As observed by Pickett's group, the reduction potential gap could be correlated with the spin density distribution derived from DFT analysis: while in μ-H species, the spin density is delocalized on both Fe atoms,[78] in t-H congener is localized on the Fe atom with no hydride.[77] Finally, Zaffaroni et al. reported that the protonation of $[Fe_2(adt)(CO)_2(PMe_3)_4]$ yields only t-H as a result. Owing to the presence of the nitrogen atom on adt as a relay group, the protonation is not only regioselective but also reversible.[79] These species however showed no tendency to eliminate H_2,[79] and this stability is consistent with the idea that the production and the oxidation of H_2 are coupled to electron-transfer reactions.[80]

Aimed at working out the issue *stable/less reactive μ-H vs unstable/reactive t-H*, one strategy has been adopted to date: the search for stable t-Hs, or at least able to persist in solution long enough to be reduced (and protonated again). Nonetheless, also because of the aforementioned evergrowing interest toward hydrogenases,[33, 71, 51] it may be worth exploring new strategies. Along this way, it can be remarked that a hydride in the bridging position between Ni and Fe atoms has been unequivocally identified by different spectroscopy techniques in [NiFe]-hydrogenase, and it has been demonstrated that is a catalytically active intermediate.[81] Even outside the hydrogenase family, other outstanding examples of bioorganometallic clusters forming μ-hydrides

$$\begin{tikzcd}
ox_{gas} \arrow[d, "\Delta G^{o}_{s}(ox)"'] & + & e^{-}_{gas} \arrow[r, "\Delta G^{o\,,\,redox}_{gas}"] & red_{gas} \arrow[d, "\Delta G^{o}_{s}(red)"] \\
ox_{solv} & + & e^{-}_{solv} \arrow[r, "\Delta G^{o\,,\,redox}_{solv}"'] & red_{solv}
\end{tikzcd}$$

Figure 2.1.7: Born-Haber cycle used for the DFT evaluation of redox potentials

are known, such as FeMo-co embedded in the nitrogenase active site.[82] This would suggest that, in nature, there is no intrinsic disadvantage for catalytic processes to pass through metal—metal bridging hydride forms. Thus, we have searched with DFT methods for the combination of ligands that afforded species with $\Delta E°_{\text{t-H}-\mu\text{-H}} < 0$, thus indicating a higher (or at least as high as) redox reactivity of µ-H versus t-H. This would allow us to overturn the paradigm, valid certainly in the [FeFe]-hydrogenase modeling, that depicts µ-H as persistent-less reactive, whereas t-H is labile-reactive.

2.1.2.2 Methods

All DFT calculations have been carried out with the TURBOMOLE suite of programs,[57] applying the resolution of identity technique,[58] using the all-electron TZVP basis set[59] and the BP86 pure functional.[60, 61] This level of theory has been shown to be suitable and reliable for the investigation of [FeFe]-hydrogenase models.[62, 63, 48, 56, 64] Aiming at computing theoretical values of standard reduction potentials (E°) that best reproduced those experimentally observed, we initially adopted a protocol based on the Born—Haber cycle.[83, 84] Free-energy changes associated with the half-reactions are shown in Figure 2.1.7. The standard Gibbs free energy is calculated as follows:

$$\Delta G^{o\,,\,redox}_{solv} = \Delta G^{o\,,\,redox}_{gas} + \Delta G^{o}_{s}(red) - \Delta G^{o}_{s}(ox) \qquad (2.1.1)$$

The standard one-electron redox potential E° is calculated by eq. 2.1.2:

$$\Delta G^{\mathrm{o}\,,\,redox}_{solv} = -FE^{o} \qquad (2.1.2)$$

where ΔG is calculated as in eq. 2.1.1 and F is the Faraday constant, 23.061 kcal mol^{-1} V^{-1}. In order to enclude entropy, zero-point energy (ZPE) and

$\Delta E^{\circ}{}_{tH-\mu H}$					
ΔE	ΔE_{solv}	$\Delta E_{solv(SCF)}$	ΔG	ΔG_{solv}	$\Delta E_{solv(SCF)}$
179	216	180	136	172	144

Table 2.3: $\Delta E^{\circ}{}_{tH-\mu H}$ values (in mV) calculated at different levels of theory for $[HFe_2(pdt)(CO)_2(dppv)_2]^+$

enthalpy contribution into the self-consisted field energy (E_{SCF}), free energy values (G) were calculated considering three different contributions to the total partition function ($q_{translational}$, $q_{rotational}$ and $q_{vibrational}$).[19] The T and P values were generally set at 298.15 K and 1 bar respectively (other T values have been used in specific cases in which the experimental conditions were explicitly different). Because of the high values of the SCF wavenumbers arising from the harmonic approximation, a scaling factor of 0.9914 was applied.[57]

The solvent (CH_2Cl_2) polarizing effect has been modeled according to the COSMO approach,[85, 86] by considering a polarizable continuum medium characterized by a ε=9.01. Because the experimental value is known for the gap of the standard reduction potential ($\Delta E^{\circ}{}_{tH-\mu H}$) associated with the aforementioned species $[HFe_2(pdt)(CO)_2(dppv)_2]^+$,[76] we have tested which method provided the best match of theory *versus* experiment, among the following: ΔE (gas-phase condition), ΔE_{solv} (struture optimization in COSMO CH_2Cl_2), $\Delta E_{solv(SCF)}$ (gas-phase value + single point COSMO CH_2Cl_2), ΔG (ΔE + thermal/entropic corrections), ΔG_{solv} (ΔE_{solv} + thermal/entropic correnctions) and ΔG_{solv} ($\Delta E_{solv(SCF)}$ + thermal/entropic correnctions). All results for ΔE° calculations are reported in Table 2.3.

$\Delta G_{solv(SCF)}$ represents in theory the formally correct approach because it includes entropic and thermal contributions and solvent effects as well. However, in the case under investigation, $\Delta E_{solv(SCF)}$ values afforded the closest match of theory *versus* experiment concerning $\Delta E^{\circ}{}_{tH-\mu H}$ (exp value, 200 mV;[76] computed value arising from $\Delta E_{solv(SCF)}$, 216 mV; value from ΔG_{solv}, 172 mV). Besides reproducing $\Delta E^{\circ}{}_{tH-\mu H}$, our computational model allowed us to obtain a fine match (max error = 0.14 V) of theory *versus* experiment of redox potentials referenced to electrodes employed experimentally (such as Fc^+/Fc).[84] The better performance of ΔE *versus* ΔG in the present case is likely due to a cancelation error effect arising from approximations normally used to turn ΔE into ΔG (i.e., ideal gas approximation, separation into components of the total energy, harmonic approximation, etc.).

2.1.2.3 Results and discussion

To investigate the possibility of designing bridging hydrides that show more favorable reduction potentials than terminally coordinated congeners and implicitly unveil which factors are more relevant in making one electron reduction of t-H generally easier than that of bridging hydrides,[76, 77] we studied a series of [FeFe]-hydrogenases bioinspired derivatives. We first varied all carbonyl and cyanides ligands (L) but the apical one at [Fe_p] and the pdt chelate. Then, since the methylthiolate ligand has never been employed in biomimetic compounds, we investigated the effect of its replacement with other ligands (Y) employed in inorganic synthesis, such as CO, PMe_3, and N-heterocyclic carbene (NHC), with this latter already proven to be a good donor, similarly to the biological cyanide ligand. Finally, also the effects of switching (XDT) chelate to different forms have been investigated.

Effects of variations of L ligands

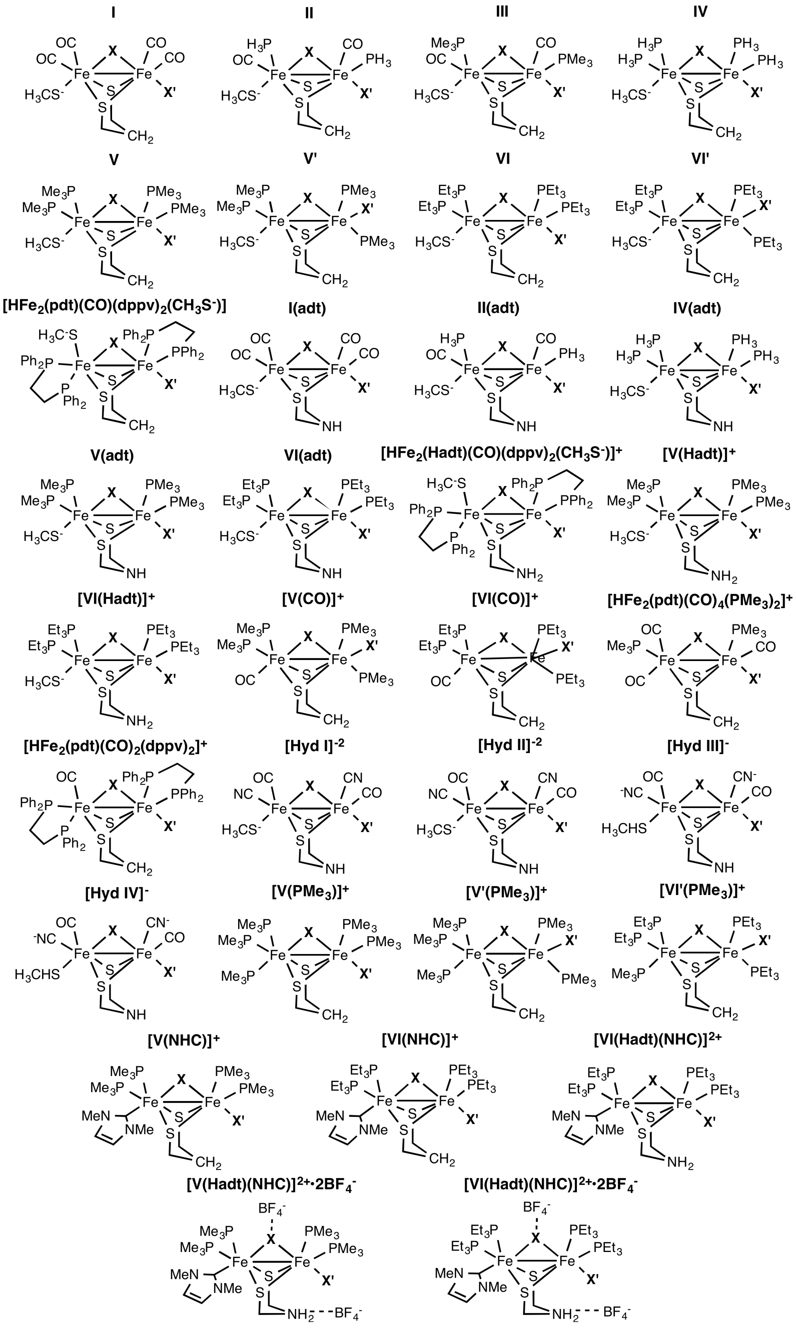
Because cyanides have been seldom used as iron ligands in the synthetic modeling of hydrogenases, due to their undesired proton affinity,[51] carbonyls and different types of phosphines[87, 88] have been used here as ligands of the [2Fe2S] centre. Starting from the atomic coordinates of the PDB structure (PDB ID:1HFE), canonical H-cluster ligands were modified to obtain a series of compounds: $[HFe_2(pdt)(PR_3)_n(CO_2)_{5-n}CH_3S^-]$ (with n = 0, 2, 4; R = H, Me, Et; Figure 2.1.8). The whole set of complexes is shown in Figure 2.1.8, with X/X' representing H and CO ligands switching relative positions in bridging (X=H^-, X'=CO) *versus* terminal hydrides (X=CO, X'=H^-). The H-cluster models investigated in the present study are: **[Hyd I]**$^{2-}$ = $[HFe_2(pdt)(CO)_3(CN^-)_2(CH_3S^-)]^{2-}$, **[Hyd II]**$^{2-}$ = $[HFe_2(adt)(CO)3(CN)_2(CH_3S^-)]^{2-}$, **[Hyd III]**$^-$ = $[HFe_2(pdt)(CO)3(CN)_2(CH_3SH)]^-$, and **[Hyd IV]**$^-$ = $[HFe_2(adt)(CO)3(CN^-)_2(CH_3SH)]^-$. Only the lowest-energy stereoisomer (related to possible different rotamers at the [Fe_d]$(X)_3$ group) of each species is shown in the picture (Figure 2.1.8). In **V´**, **VI´**, **[V(CO)]**$^+$, **[VI(CO)]**$^+$, **[V´(PMe$_3$)]**$^+$, and **[VI´(PMe$_3$)]**$^+$ species, the correct t-H disposition is to be obtained by rotating the [Fe_d]$(X)_3$ group by 120°. Upon reduction, the diiron core of μ-H and t-H is formally reduced from [$Fe^{II}Fe^{II}$]

compound	$E^\circ_{\mu H}$	E°_{tH}	$\Delta E^\circ_{tH-\mu H}$
I	-1.782	-1.589	193
II	-2.415	-2.230	185
III	-2.677	-2.478	199
IV	-2.990	-2.836	154
V	-3.451	-3.175	276
V'	-3.249	-3.175	74
VI	-3.185	-3.198	-13
VI'	-2.648	-3.198	-550

Table 2.4: Computed redox potentials (*vs* Fc^+/Fc) of μ-H and t-H for the series $[HFe_2(pdt)(PR_3)_n(CO)_{5-n}CH_3S^-]$

to $[Fe^{II}Fe^{I}]$. Here, the values for $E^\circ_{\mu H}$ and E°_{tH} and $\Delta E^\circ_{tH-\mu H}$ have been referenced to the absolute redox potential of the Fc^+/Fc couple, commonly employed in electrochemical investigations related to [FeFe]-hydrogenases. The aim of systematically very the number and type of phosphines is to evaluate the role of their electron content and the steric bulk in modulating $\Delta E^\circ_{tH-\mu H}$. In addition, because the complexes in the model set contain an apical methylthiolate at $[Fe_p]$, this analysis has been done with the goal of modeling the role of the biogenic cysteine.

The results in Table 2.4 show that t-H is more easily reduced than μ-H, in all species considered (except **VI** and **VI´**), which is generally in line with the experimental data related to the biomimetic models investigated to date.[76] As expected, increasing the electronic density on the [2Fe] core of the various species implies more negative E°. Nonetheless, increasing the number of phosphines coordinated to the [2Fe] core does not appear to significantly affect the reduction potential gap, showing rather an irregular trend, on going from **I** to **V**, with the last one featuring the clearest preference for t-H reduction. Unexpectedly, replacing PMe_3 with PEt_3 causes E° to become approximately equivalent in t-H *versus* μ-H ($\Delta E^\circ_{tH-\mu H} = -13$ mV). Moreover, **VI** is also the only species showing a bridging form less stable than the terminal one, at both the $[Fe^{II}Fe^{II}]$-$[H^-]$ (by 4.4 kcal mol^{-1}) and $[Fe^{II}Fe^{I}]$-$[H^-]$ (by 4.1 kcal



I **II** **III** **IV**

V **V'** **VI** **VI'**

[HFe$_2$(pdt)(CO)(dppv)$_2$(CH$_3$S$^-$)] **I(adt)** **II(adt)** **IV(adt)**

V(adt) **VI(adt)** **[HFe$_2$(Hadt)(CO)(dppv)$_2$(CH$_3$S$^-$)]$^+$** **[V(Hadt)]$^+$**

[VI(Hadt)]$^+$ **[V(CO)]$^+$** **[VI(CO)]$^+$** **[HFe$_2$(pdt)(CO)$_4$(PMe$_3$)$_2$]$^+$**

[HFe$_2$(pdt)(CO)$_2$(dppv)$_2$]$^+$ **[Hyd I]$^{-2}$** **[Hyd II]$^{-2}$** **[Hyd III]$^-$**

[Hyd IV]$^-$ **[V(PMe$_3$)]$^+$** **[V'(PMe$_3$)]$^+$** **[VI'(PMe$_3$)]$^+$**

[V(NHC)]$^+$ **[VI(NHC)]$^+$** **[VI(Hadt)(NHC)]$^{2+}$**

[V(Hadt)(NHC)]$^{2+}\cdot$2BF$_4^-$ **[VI(Hadt)(NHC)]$^{2+}\cdot$2BF$_4^-$**

Figure 2.1.8: Full set of models investigated in the present study.

mol^{-1}) states. As for the **V´** and **VI´** μ-H species, they are isomers (with basal–apical disposition of PX_3 ligands at [Fe_p]) of **V** and **VI**, respectively. Reduction of these isomers has been investigated because they are very close in energy (within the DFT accuracy value) to the dibasal forms **V** and **VI**. Thus, both **V** and **VI** should be considered as two-component systems, denoted as **V**/**V´** and **VI**/**VI´**, given a likely coexistence of both isomers in solution. Like **VI**, also **VI´** features a t-H more stable than μ-H by more than 2 kcal mol^{-1} at the [$Fe^{II}Fe^{II}$]-[H^-] state, whereas **V**/**V´** shows a preference for μ-H. This result is serendipitous, although particularly relevant, because stabilization of terminal *versus* bridging hydrides has been and still is to be considered a challenge from a synthetic standpoint. Investigation of all of the reduced μ-H structures reveals that only **VI** features one [Fe_p]-H^- bond that is completely cleaved, to such an extent that H^- coordination becomes terminal. In light of the reported higher reactivity toward protons of t-H form [74] and because a second protonation is a requisite for the HER occurrence, the characteristics identified in **VI** could unravel interesting perspectives. In particular, because $\Delta E°_{tH-\mu H}$ associated with **VI** is near zero, whereas it is clearly > 0 for the very similar **V**, a proper tailoring of the second coordination sphere of the [2Fe] core could allow to tune such redox parameters. Curiously, however, the narrowing (and the sign inversion) of $\Delta E°_{tH-\mu H}$ in **VI** is concomitant with the presence of a more stable t-H than μ-H so as to indicate the "classical" strategy and that pursued herein could be related. It could be argued that PMe_3 and PEt_3 have so similar basic properties that the discrepancy that emerged between **V** and **VI** should be ascribed to steric repulsions possibly induced by the different bulk of R groups. It has to be noticed that going from electron-poorer to electron-richer compounds, the number/type of P ligands does not clearly affect the E° gap, whose sign, yet, is even reverted in the presence of bulky Et groups on 4-P species. This would suggest that in a highly substituted diiron core, the creation of a conditions where bulky phospines induce an intramolecular steric effects can play a role and have an evident impact on $\Delta E°_{tH-\mu H}$ modulation. The nonlinear trend outlined in Table 2.4 for $\Delta E°_{tH-\mu H}$ shows that sign inversion occurs in an unpredictable way. This supports the fact that the number of σ-donors on the diiron derivative is not intrinsically a determinant factor, but it becomes a crucial prerequisite as it allows some peculiar effects to unveil their role on $\Delta E°_{tH-\mu H}$ variation. Even more surprisingly, not only the number of σ-donors but also the donor strength is not per se sufficient to rationalize the observed picture. As previously extensively described,[89, 90, 91, 92] the PEt_3 group is a stronger donor ligand than PMe_3 is, even though other properties such as the pK_a's of PMe_3 and PEt_3 are quite similar (8.65 for PMe_3 *vs* 8.69 for PMe_3).[93] In addition, $E°_{\mu H}$ of **V** is more negative than that of **VI** and **VI´**, whereas the reversal is observed in t-H congeners of the same species. All such apparen anomalies must therefore be explained, invoking steric factors. The last statement is even reinforced considering that [Fe]-P(Et_3) (2.392 Å, avg) bond distances are generally longer than those of [Fe]-P(Me_3) (2.327 Å, avg). This is another counterintuitive observation with respect to the outcome that had been figured out based on the stronger donor properties of PEt_3 *versus* PMe_3.[90, 91, 92] To corroborate our hypothesis and to pinpoint the intimate reason underlying the **V** *versus* **VI** reductive behavior, a series of structural parameters have been analyzed in the **I**—**VI** set. Analysis of the data in Table 2.5 will be crucial to evaluating if and how a given structure is affected by intramolecular repulsive strain, that is the factor that will be shown to be decisive to explain our results on $\Delta E°_{tH-\mu H}$.

The interatomic distances shown in Table 2.5 aim at monitoring structural variations upon reduction in the molecular region binding the hydride, throughout the **I**—**VI** set. This allows us to gain information on both the strength of the [Fe]-[Fe] interaction and the symmetry degree of the hydride coordination to iron (terminal *vs* bridging character), upon going from $[Fe^{II}Fe^{II}]$ to $[Fe^{II}Fe^{I}]$ states.

From the data it is evident that **VI** has a peculiar feature, unique in the whole set: upon reduction, its μ-H species undergoes a complete breaking of the [Fe]-[Fe] bond with concomitant generation of relatively separated Fe subunits, the first containing a five-coordinated $[Fe_d]$ and the second including a six-coordinated $[Fe_p]$. The last one is coordinated by hydride in a completely terminal fashion. **V** instead, even if very similar to **VI**, does not behave analogously once reduced. Indeed, **V** at $[Fe^{II}Fe^{I}]$ redox state displays elongation of both [Fe]-[Fe] and $[Fe_d]$—H distances, but these are definitely moderate if quantitatively compared to those observed in **VI**. Such a different behavior in so similar derivatives occurring for only one of the two isomeric forms could help to explain the drop of $\Delta E°_{tH-\mu H}$ observed in PEt_3 *versus* PMe_3 species. Otherwise stated, the cause of the structural differences for **V** *versus* **VI** in their reduced state could be at the ground of the sign inversion of $\Delta E°_{tH-\mu H}$. Intramolecular steric repulsions are introduced by enlarging the R size of phos-

μ-H						
	$[\mathrm{Fe^{II}Fe^{II}}]$			$[\mathrm{Fe^{II}Fe^{I}}]$		
	[Fe]-[Fe]	$[\mathrm{Fe_p}]$-H	$[\mathrm{Fe_d}]$-H	[Fe]-[Fe]	$[\mathrm{Fe_p}]$-H	$[\mathrm{Fe_d}]$-H
I	2.600	1.644	1.721	2.779	1.593	1.854
II	2.598	1.641	1.739	2.690	1.626	1.765
III	2.615	1.640	1.750	2.799	1.589	1.893
IV	2.592	1.689	1.789	2.678	1.655	1.844
V	2.719	1.689	1.790	2.831	1.656	1.844
VI	2.837	1.751	1.823	3.419	1.545	2.963

t-H						
	$[\mathrm{Fe^{II}Fe^{II}}]$			$[\mathrm{Fe^{II}Fe^{I}}]$		
	[Fe]-[Fe]	$[\mathrm{Fe_p}]$-C(O)	$[\mathrm{Fe_d}]$-C(CO)	[Fe]-[Fe]	$[\mathrm{Fe_p}]$-C(CO)	$[\mathrm{Fe_d}]$-C(CO)
I	2.526	2.000	1.978	2.704	1.990	1.998
II	2.505	1.976	1.953	2.661	1.960	1.973
III	2.501	1.978	1.938	2.664	1.961	1.966
IV	2.487	1.992	1.992	2.618	1.941	1.965
V	2.541	1.922	1.922	2.672	1.941	1.966
VI	2.573	1.995	1.931	2.723	1.943	1.978

Table 2.5: [Fe-Fe] and other selected distances (Å) in the μ-H and t-H isomers of the **I-VI** set

phine ligands upon going from **V** to **VI**, with the consequence of strain production (or destabilization) in [$Fe^{II}Fe^{II}$] redox state of **VI**, which is unique in the **I**–**VI** set of compounds. The destabilization affects both μ-H and t-H of **VI**, but reduction can alleviate repulsive interactions only in the bridging isomer but not in t-H, as emerged from a comparison of the structural parameters of **V** *versus* **VI**. Thus, it becomes of primary importance to understand why reduction can separate the two Fe subunits of μ-H but not those of t-H in the case of **VI**. Some features being present exclusively in t-H may prevent the strain from being relaxed upon reduction. Indeed, all terminal-bound hydrides have a CO in the bridging position between the two Fe ions. Furthermore, the irons-bridging CO is *trans*-oriented simultaneously to two strong electron donors (H^- and CH_3S^-): it is therefore predictable that, specifically in electron-rich derivatives, CO receives a large amount of electron density through a double π- back-donation from both Fe atoms. This implies the presence of two strong [Fe_p]-C(O)-[Fe_d] bonds (each with partial double-bond character), which can prevent dissociation of the two Fe subunits upon reduction. Bridging hydrides cannot show similar behavior because of the absence of π-acceptor properties by the H ligand. This results in μ-hydrido isomers of species affected by intramolecular destabilization (exactly as **VI**) being free to alleviate the strain thanks to reduction. The observation of key distances associated with **V** (Table 2.5) brings to light that this derivative must not be affected by intramolecular strain, which implies that its reductive behavior should not deviate significantly from the other less substituted derivatives of Table 2.4. This indirectly indicates that, without steric factors influencing selectively the two hydrido forms of a given species, the observation of a more favorable redox potential associated with t-H must be governed by pure electronic factors. As aforementioned, the DFT isomer speciation showed (limited to **V** and **VI** in their [$Fe^{II}Fe^{II}$] state) that the two possible ligand dispositions at [Fe_d] ($P_{dibasal}$ and $P_{apical-basal}$) are so close in energy that DFT accuracy does not allow an unequivocal resolution of their relative stability. Therefore, the two stereoisomers should be fairly predicted to coexist in a thermodynamically equilibrated solution. Redox potentials have therefore been calculated also for **V´** and **VI´**. Interestingly, reduction causes this time the loss of the apical PEt_3 from [Fe_d] coordination in the sole instance of **VI´**, while **V´** [$Fe^{II}Fe^{I}$] was still a bis-six-coordinated Fe species. The [Fe_d]–PEt_3 bond breaking upon reduction leads to a larger release of molecular strain compared to that observed for reduction of the dibasal isomer. Thus, dissociating a P ligand from Fe represents the other way that μ-Hs of strained/unstable species have for relaxing accumulated repulsions. Of course, because **VI** and **VI´** share the same t-H structure, the rationale that justifies why the strain release occurs only in μ-Hs is the same as that outlined previously. These results support quite incontrovertibly the idea that a proper modulation of the intramolecular sterics can influence the $\Delta E°_{tH-\mu H}$ gap. Frontier molecular orbital (FMO) analysis has been then performed to pinpoint which electronic properties are mostly influenced by the structural difference in **V** *versus* **VI**. From inspection of the singly occupied molecular orbital (SOMO) in reduced states of **V** *versus* **VI**, shown in Figure 2.1.9, it is evident that the molecular strain affecting **VI** and (but not **V**) finds release upon reduction in μ-H, whereas in t-H, there is actually a strain retention. A comparison of E_{SOMO} values shows that there is complete consistence with the more favorable reduction of t-H *versus* μ-H in **V** ($\Delta E°_{tH-\mu H} > 0$; Table 2.4) and simultaneously with the reverse behavior in **VI** ($\Delta E°_{tH-\mu H} < 0$). Also some differences are detectable in the atomic orbital composition (with consequential unequal distribution) of the two SOMOs in μ-hydrides of **V** *versus* **VI**, whereas such a parameter is practically unchanged when looking at the SOMO of t-Hs of the same species.

Effects of variations of Y ligand

Heretofore, have been considered complexes featuring the CH_3S^- ligand, a mimic of the cysteinyl sulfur bridge that links the [4Fe4S] cluster to the $[2Fe_H]$ center.[33, 34, 39, 94] Nonetheless, has been already mentioned that CH_3S^- is not normally employed in the synthetic hydrogenases' chemistry.[95] Thus, has been set out to investigate the effect of replacing it with other ligands more routinely used synthetically, such as neutrally charged CO, PMe_3, and also NHC. The results showing E°, $\Delta E°_{tH-\mu H}$, and all of those structural parameters that are useful to evaluating the strain release upon reduction of the various species are reported in Tables 2.6 and 2.7.

First, it is evident that the **V**-like *versus* **VI**-like comparison confirms the essential role of intramolecular repulsions and their release upon reduction for justifying the $\Delta E°_{tH-\mu H}$ variation. The presence of the CO group in place of the methylthiol group makes $\Delta E°_{tH-\mu H}$ significantly positive, indicative of μ-H reduction clearly disfavored *versus* t-H reduction. The importance of the

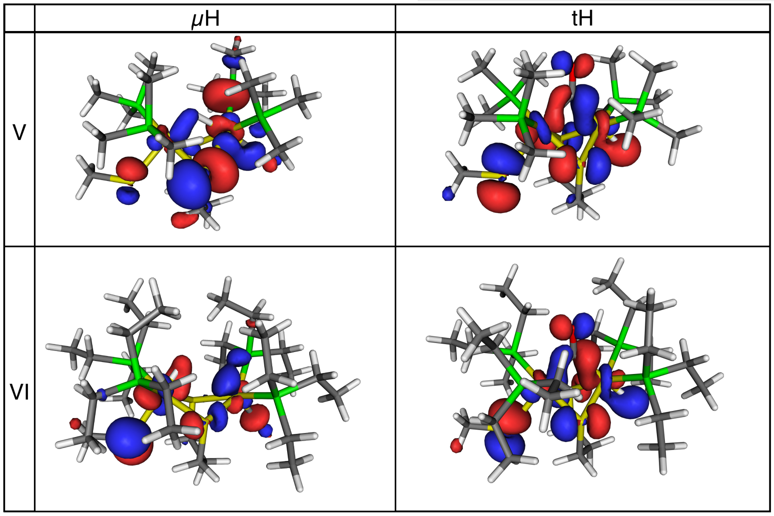


Figure 2.1.9: E_{SOMO}(eV) and electronic distribution in reduced states of both t-H and μ-H for derivatives **V** and **VI**. The isosurface cutoff value is 0.03 atomic units.

compound	$E^{\circ}_{\mu H}$	E°_{tH}	$\Delta E^{\circ}_{tH-\mu H}$
[V(CO)]$^+$	-2.201	-2.022	179
[VI(CO)]$^+$ *	-2.649	-1.944	705
[V(PMe_3)]$^+$	-2.645	-2.353	292
[V'(PMe_3)]$^+$	-2.467	-2.354	114
[VI(PMe_3)]$^+$ *	-1.467	-2.042	-566
[V(NHC)]$^+$	-2.840	-2.626	214
[VI(NHC)]$^+$	-2.523	-2.568	-45

Table 2.6: Computed redox potentials (*vs* Fc^+/Fc) of μ-H and t-H for the series $[HFe_2(pdt)(PR_3)_n(CO)_{5-n}CH_3Y]$ (X=Me in **V/V'** and X=Et in **VI/VI'**, Y=CO, $PMe_3$and NHC). In complexes designated with an asterisk, the reduced form of μ-H loses an apical phosphine, originally *trans*-oriented to hydride, whereas **[V(NHC)]$^+$** is the only 4-PEt_3 system in which the strain release occurs through breaking of Fe—Fe and one Fe—H bonds.

	μ-H			
	$[Fe^{II}Fe^{II}]$		$[Fe^{II}Fe^{I}]$	
	[Fe]-[Fe]	$[Fe_d]$-H	[Fe]-[Fe]	$[Fe_d]$-H
[V(CO)]$^+$	2.648	1.662	2.792	1.680
[VI(CO)]$^+$	2.689	1.675	2.692	1.695
[V(PMe$_3$)]$^+$	2.765	1.748	2.932	1.776
[V'(PMe$_3$)]$^+$	2.718	1.702	2.885	1.802
[VI(PMe$_3$)]$^+$	2.779	1.710	2.719	1.721
[V(NHC)]$^+$	2.711	1.702	2.873	1.846
[VI(NHC)]$^+$	2.881	1.790	3.527	3.011

	t-H			
	$[Fe^{II}Fe^{II}]$		$[Fe^{II}Fe^{I}]$	
	[Fe]-[Fe]	$[Fe_p]$-C(CO)	[Fe]-[Fe]	$[Fe_p]$-C(CO)
[V(CO)]$^+$	2.641	1.529	3.093	2.999
[VI(CO)]$^+$ *	2.674	2.513	3.179	2.989
[V(PMe$_3$)]$^+$	2.772	2.747	3.268	3.279
[V'(PMe$_3$)]$^+$	2.772	2.747	3.268	3.279
[VI(PMe$_3$)]$^+$ *	2.920	2.921	3.379	3.297
[V(NHC)]$^+$	2.768	2.679	3.214	3.144
[VI(NHC)]$^+$	2.875	2.771	3.329	3.181

Table 2.7: [Fe-Fe] and other selected distances (Å) in the μ-H and t-H isomers of species in Table 2.6. In complexes designated with an asterisk, reduced μ-H loses an apical phospine *trans*-oriented to hydride, whereas **[V(NHC)]$^+$** is the only 4-PEt$_3$ system in which the strain release occurs through breaking of Fe—Fe and one Fe—H bonds.

simultaneous presence of two σ-donors *trans*-oriented to bridging CO in terminal hydrides has already been invoked as the reason that prevents strained t-H (of the **VI/VI´** system) from being energy-relaxed. Therefore, t-H isomer of **[VI(CO)]**$^+$ species can relax, as well as µ-H, thus not showing that different behavior upon reduction of t-H *versus* µ-H, which is necessary to observe $\Delta \mathrm{E}^{\circ}{}_{\mathrm{tH}-\mu\mathrm{H}}$ inversion. Indeed, the true reason for which the oxidized **[VI(CO)]**$^+$-t-H is strained, despite the presence of two acceptors (with the CO replacing CH_3S^- and µ-CO), whereas, for example, **[VI´(PMe$_3$)]**$^+$ is not strained is rather subtle and is more complicated than depicted here. It requires invoking electronic effects, which have been discussed in the following paragraphs. It was expected that substitution of CH_3S^- by PMe_3(**[VI´(PMe$_3$)]**$^+$) would lead to outcomes analogous to those of **VI/VI´**. These findings can be explained by the σ-donor nature of PMe_3, which, although less strong compared to the thiolate, is well-known. In this peculiar case, it has been studied only one stereoisomer at [Fe$_\mathrm{d}$] because the other is very unstable (data not shown). **[VI´(PMe$_3$)]**$^+$ actually shows strongly negative $\Delta \mathrm{E}^{\circ}{}_{\mathrm{tH}-\mu\mathrm{H}}$ (see Table 2.6), but the reason underlying this result requires application of the *strain-release model* in a slightly different way. Because the apical PMe_3 is a weaker donor than thiolate, it cannot trigger the formation of a [Fe$_\mathrm{p}$]—C(O) bond (2.921 Å) as strong as in the case of CH_3S^--based species **VI/VI´** (1.995 Å), essential for preventing relaxation of molecular strain in terminal hydrides. This entails that t-H of **[VI´(PMe$_3$)]**$^+$ is not strained at all already at the oxidized level (see Figure 2.1.10), as is clearly evidenced from the structural parameter analysis presented in Table 2.7.

$\Delta \mathrm{E}^{\circ}{}_{\mathrm{tH}-\mu\mathrm{H}}$ values and structural parameters associated with **[VI(NHC)]**$^+$ qualitatively confirm all results observed for **[VI´(PMe$_3$)]**$^+$. The quantitative difference ($\Delta \mathrm{E}^{\circ}{}_{\mathrm{tH}-\mu\mathrm{H}}$ of the PMe_3 species is clearly more negative than that in the NHC one) is explained by the fact that the **[VI´(PMe$_3$)]**$^+$ µ-H loses a phosphine upon reduction, while the **[VI(NHC)]**$^+$ µ-H displays [Fe]—[Fe] and [Fe$_\mathrm{d}$]—H bond breaking. This last reductive behavior has less impact on the E° gap (see also **VI** *vs* **VI´**) from a quantitative standpoint. It may be argued at this point that also **[VI'(CO)]**$^+$ should feature a [Fe$^{\mathrm{II}}$Fe$^{\mathrm{II}}$] t-H repulsively unstrained, just like **[VI´(PMe$_3$)]**$^+$ and **[VI(NHC)]**$^+$. The apical CO is actually expected to induce an even weaker [Fe$_\mathrm{p}$]—C(O) bond *trans* to it, compared to those characterizing **[VI´(PMe$_3$)]**$^+$ and **[VI(NHC)]**$^+$. As can be viewed in Table 2.7 (for example, focusing on the [Fe]—[Fe] bond),

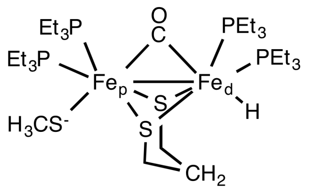
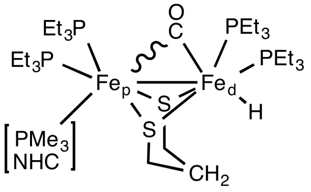
t-H $Fe^{II}Fe^{II}$	
VI	**$[VI'(PMe_3)]^+/[VI(NHC)]^+$**
	

Figure 2.1.10: Neutrally charged apical donors implying a weakening of the $[Fe_p]$—C(O) bond in the cationic diferrous state of t-H: this causes **[VI´(PMe$_3$)]**$^+$ and **[VI(NHC)]**$^+$ to be unstrained structures, unlike **VI** (bearing apical CH_3S).

[VI’(CO)]$^+$ features structural parameters that are indicative of a strained oxidized state, at least more than the corresponding state in **[VI´(PMe$_3$)]**$^+$ and **[VI(NHC)]**$^+$. That is due to a couple of factors, the first being the fact that the most stable rotameric disposition of **[VI’(CO)]**$^+$ at $[Fe_p]$ is not $(CO)_{ap}$—Fe—$(PEt_3)_{2ba-ba}$ but instead $(PEt_3)_{ap}$—Fe—$(CO)_{ba}(PEt_3)_{ba}$. Such a result is probably due to the same aforementioned reason: two mutually *trans*-oriented CO’s are less stable than that in a P—Fe—(CO) disposition. Thus, oxidized t-H-**[VI’(CO)]**$^+$ has actually a donor (P) ligand in the apical position at $[Fe_p]$, just like **[VI´(PMe$_3$)]**$^+$ and **[VI(NHC)]**$^+$, so that the issue of the different strain (in the former vs the two latter species) remains unclear, unless electronic factors are invoked. To shed light on that, one has to consider that t-H-**[VI’(CO)]**$^+$ has a less electron-rich $[2Fe_H]$ core than that of the other two species because of the presence of CO *versus* NHC and PMe_3. This means that in t-H-**[VI’(CO)]**$^+$ more electron density is drawn from the $[2Fe_H]$ core compared to t-H-**[VI´(PMe$_3$)]**$^+$ and t-H-**[VI(NHC)]**$^+$ (see Figure 2.1.11).

Although the 18-electron rule formally requires the presence of the [Fe]—[Fe] bond in the investigated species, it is yet conceivable that, in very electron-rich di-iron cores such as those in t-H-**[VI´(PMe$_3$)]**$^+$ and t-H-**[VI(NHC)]**$^+$, each Fe ion is already electronically saturated, so that the [Fe]—[Fe] formation is less mandatory than that in t-H-**[VI’(CO)]**$^+$. On a more general level, it can thus be confidently concluded that what ultimately determines the sign of $\Delta E°_{tH-\mu H}$

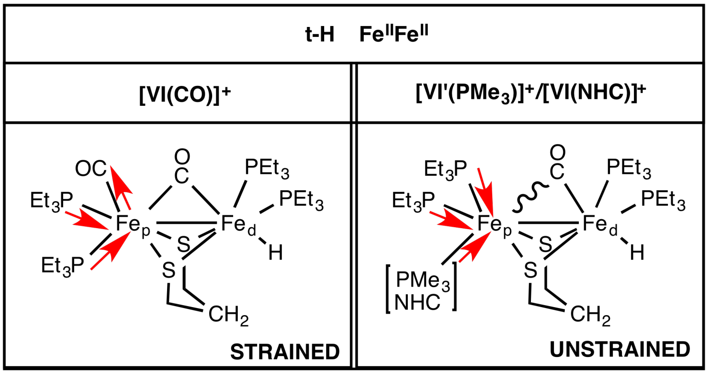


Figure 2.1.11: Schematic illustration of electronic factors underlying the presence of strain in t-H-**[VI'(CO)]**$^+$ and the absence of strain in t-H-**[VI´(PMe$_3$)]**$^+$ and t-H-**[VI(NHC)]**$^+$

is not necessarily the presence of factors keeping strained t-Hs from relaxing, whereas corresponding μ-Hs are able to do that (as illustrated in the **VI**/**VI´** case). Indeed, the crucial point is the different reductive behavior of t-H *versus* μ-H of a given species, in whatever fashion it may occur. It is relevant that the model that has been devised to monitor the presence/absence of intramolecular strain remains valid in all of the instances investigated, with the only variation residing in the manner of application of the model itself. Identical behavior implies that t-Hs are reduced at potentials less negative than those of μ-Hs, whereas different behavior implies the opposite outcome. It has been stressed that herein the *identical/different* couple refers to how/if the t-H *versus* μ-H of a particular instance changes its features in terms of the repulsive strain model that has been set up and which could be coded as follows. Let 'S' stand for strained and 'US' for unstrained molecule: an identical reductive behavior t-H *versus* μ-H will be coded either by 'S/S *vs* S/S' or by 'US/US *vs* US/US' or finally by 'S/US *vs* S/US'. Any other combination of characters associated with the strain model will provide a different reductive behavior in the t-H/μ-H couple. This rationale can be schematically shown in Table 2.8, from which it is clear how a given pattern (the array of S/US characters) associated with the presence/absence of repulsive strain in both t-H and μ-H of each derivative entails a positive or negative sign of $\Delta E°_{\mathrm{tH}-\mu\mathrm{H}}$.

In light of the literature reporting that NHC and PMe$_3$ have been previously incorporated in diiron dithiolates modeling [FeFe]-hydrogenases,[78, 79, 96]

	μ-H		t-H		
compound	[$Fe^{II}Fe^{II}$]	[$Fe^{II}Fe^{I}$]	[$Fe^{II}Fe^{II}$]	[$Fe^{II}Fe^{I}$]	$\Delta E°_{tH-\mu H}$
V/V'	US	US	US	US	> 0
VI/VI'	S	US	S	S	< 0
[V(CO)]$^+$	US	US	US	US	> 0
[VI(CO)]$^+$	S	US	S	US	> 0
[V/V'(PMe$_3$)]$^+$	US	US	US	US	> 0
[VI(PMe$_3$)]$^+$	S	US	US	US	< 0
[V(NHC)]$^+$	US	US	US	US	> 0
[VI(NHC)]$^+$	S	US	US	US	< 0

Table 2.8: Strain pattern related to some key species described above.

derivatives equipped with such ligands in the Fe apical position **[VI´(PMe$_3$)]$^+$** and **[VI(NHC)]$^+$** could be potential candidates for the synthesis of new biomimetic compounds.

Effects of protonation of CH_3S^- ligand

Further, it has been investigated the possibility of altering $\Delta E°_{tH-\mu H}$ by decreasing the electron-donor character of the CH_3S^- group by protonating the S atom (which could mimic the electron-attracting character of the CO group and the presence of the cubane cluster). Such an effect has been tested on a set of computational models of the active site of [FeFe]-hydrogenase (here named **[Hyd X]**). Both dithiolate straps adt and pdt have been considered, a comparison anyway useful for the purposes of the present study, although the issue of the real identity of the X element in $-SCH_2XCH_2S-$ has been finally resolved in favor of nitrogen.[97] The presence of CH_3SH in place of the anionic CH_3S^- leads to an increase of $\Delta E°_{tH-\mu H}$ by more than 300 mV in the model featuring pdt, while the augment is more limited in the case of the adt form (Table 2.9).

An easier reduction of the t-H form fits well with the literature, which describes a scenario in which enzyme catalysis is characterized, among other factors, also by kinetic stabilization of the hydride in the terminal position at [Fe_d]. The t-H-to-μ-H rearrangement, active in all biomimetic compounds, is hindered

compound	$\Delta E°_{tH-\mu H}$
[HydI]$^-$	152
[HydII]$^-$	222
[HydIII]$^-$	478
[HydIV]$^-$	307

Table 2.9: $\Delta E°_{tH-\mu H}$ associated with the [2Fe$_H$] subunit of the H-Cluster (truncated at the cysteinyl thiolate)

in the enzyme through interaction between the ligand CN$^-$ and a conserved lysine.[68, 51] In these conditions, the formation of a μ-H isomer would bring a less favored reductive event, probably impairing the catalytic efficiency.

Effect of variation od XDT chelate

The redox potentials of **[VI(NHC)]$^+$** for both hydride forms (−2.513 V for μ-H and −2.557 V for t-H vs Fc$^+$/Fc) still reveal overpotential (although less than that of CH_3S^--based species), as expected because of the large number of donor ligands coordinated to the [2Fe$_H$] center. A possible mitigation of this effect has been obtained by replacing pdt with adt. Indeed, the central amine of adt derivatives can be protonated in solutions by acids thus bearing an extra net positive charge in diiron dithiolate.[74] However, because hydride species bearing nonprotonated adt are also detected in the CH_2Cl_2 solution, $\Delta E°_{tH-\mu H}$ has been computed also for some analogues of Table 2.4 in which pdt has been replaced with neutral azadithiolate. Moreover, two derivatives experimentally synthesized, **[HFe$_2$(pdt)(CO)$_2$(dppv)$_2$]$^+$** and **[HFe$_2$(pdt)(CO)$_4$(PMe$_3$)$_2$]$^+$** [78] ($\Delta E°_{tH-\mu H} = 188$ mV), have also been considered. These last two sets of calculations have served mainly as benchmarks for the present investigation (see the Methods section for **[HFe(pdt)(CO)$_2$ (dppv)$_2$]$^+$**). The obtained values confirm the same qualitative trend as that observed for pdt-based derivatives, even if a negative $\Delta E°_{tH-\mu H}$ cannot be observed, probably because of a (N)H−[Fe] stabilizing interaction, which selectively favors t-H. When switching to protonated (Hadt) derivatives, it can be noted that **[VI(Hadt)(NHC)]$^{2+}$** shows, as expected, less negative E° values both in μ-H (−1.797 V vs Fc$^+$/Fc) and in t-H (−1.117 V vs Fc$^+$/Fc). Regrettably, yet, this result goes to the detriment of the desired $\Delta E°_{tH-\mu H}$ sign, as is

compound	$\Delta E°_{tH-\mu H}$
$[V(Hadt)]^{+}$	576
$[VI(Hadt)]^{+}$	399
$[VI(Hadt)(NHC)]^{2+}$	680
$[V(Hadt)(NHC)]^{2+}$. $2BF_4^-$	218
$[VI(Hadt)(NHC)]^{2+}$. $2BF_4^-$	84
$[HFe_2(Hadt)(CO)(dppv)_2CH_3S^-]^{+}$	455
$[HFe_2(pdt)(CO)(dppv)_2CH_3S^-]$	124

Table 2.10: Standard reduction potential gap ($\Delta E°_{tH-\mu H}$ in mV) of derivatives containing Hadt as a dithiolate strap.

observable in Table 2.10.

The reasons for that can be found in a strong intramolecular $NH(H)^{2+}$—H[Fe] interaction (dihydrogen bonding), which selectively favors t-H *versus* μ-H, a feature recently illustrated also at the solid state in **$[HFe_2(Hadt)(CO)_2(dppv)_2]^{2+}$** by Rauchfuss' group.[74] However, the strength of this interaction and also the doubly positive charge of **$[VI(Hadt)(NHC)]^{2+}$** could be overestimated by the previous calculations because of the absence of explicit counterions such as BF_4^-, which could be present near the electrode surface in buffers used in a real cyclic voltammetry experiment. In addition, it has been reported that BF_4^- ions can alter a delicately balanced equilibrium existing between N- and Fe-protonated isomers.[74] The results show that adding explicit BF_4^- ions in the simulation brings about an 8-fold decrease of $\Delta E°_{tH-\mu H}$ with respect to the isolated bis-cation (see Table 2.10), which definitively points to a strong effect of the molecular charge on the redox gap. Although marked, however, the charge effect is not sufficient in this case to observe a μ-H reduction favored over that of t-H. It is worth recalling, nonetheless, that 84 mV (1.9 kcal mol^{-1}) is a value essentially very close to the precision limit of the employed computational method, thus indicating that a species may exist with t-H and μ-H with roughly the same reduction potential. The emerging scenario reveals that the amount of negative charge that is pushed into the diiron system is able to significantly affect $\Delta E°_{tH-\mu H}$. This aspect is to be taken into appropriate consideration for fine-tuning a parameter, which could be relevant in the design of new catalysts. It becomes thus clear that a problem arises when pursuing the alternative strategy that have been herein explored, and that is thus worth recalling at this point: the design of a model species characterized by $\Delta E°_{tH-\mu H} < 0$. This result is indeed achievable but only by very electron-rich derivatives, thus obviously characterized by reduction overpotential. Any attempt that has been made to decrease E° entails a $\Delta E°_{tH-\mu H}$ shift toward more positive values. This point is confirmed also upon a comparison of **$[HFe_2(Hadt)(CO)(dppv)_2CH_3S^-]^+$** *versus* **$[HFe_2(pdt)(CO)(dppv)_2CH_3S^-]$** (see Table 2.10), in which it has been probed the effect of replacing a CO with CH_3S^- in Hadt and pdt derivatives, experimentally synthesized. CH_3S^- causes a $\Delta E°_{tH-\mu H}$ shift toward less positive values (from 216 to 124 mV) in the pdt species, which is compatible with the role that has been described above for such a ligand. The case simulation of **$[HFe_2(pdt)(CO)(dppv)_2CH_3S^-]$** is also interesting in that it allows one to call back the importance of the repulsive strain factor, as defined previously: this derivative has the same overall charge as **VI/VI´**, and all of these species share a common CH_3S^- bound at $[Fe_p]$. Despite all similarities, however, $\Delta E°_{tH-\mu H}$ is clearly positive in **$[HFe_2(pdt)(CO)(dppv)_2CH_3S^-]$** and negative in **VI** (strongly negative in **VI´**), a result that can be explained by the different intramolecular repulsive bumps that characterize these derivatives. In fact, phenyl rings of the dppv chelate can be observed to take on stabilization of T-shape orientations, at both the computational and X-ray crystallography level.[74] Also, in this case, however, upon switching to the Hadt species, a strong shift of $\Delta E°_{tH-\mu H}$ is observed, as in all preceding cases, demonstrating (once again) how an unequivocal connection exists between absolute redox potentials and their gap in t-H *versus* μ-H isomers. A compromise between the reduction overpotential and the desired $\Delta E°_{tH-\mu H}$ is therefore a probable key to design more efficient electrocatalysts. In such a sense, the tiny value computed for the redox gap value found for **$[VI(Hadt)(NHC)]^{2+}\cdot 2BF_4^-$** could be considered stimulating.

2.1.2.4 Conclusions

Increasing the number of electron donors on the protonated $[2Fe_H]$ core of [FeFe]-hydrogenases is predicted to alter $\Delta E°_{tH-\mu H}$ in a nonlinear way, up to a substitution extent at which subtle steric factors come into play entailing a surprising modulation of the redox gap. Intramolecular steric repulsions among bulky PR_3 ligands of highly substituted $(Fe_2(PR_3)_4)$ diiron dithiolates cause an unsuspected effect on $\Delta E°_{tH-\mu H}$: an inversion of its sign, the target of this present investigation. Variation upon reduction of the structural and energetic parameters shows that such a result is due to a steric strain (or destabilization), which affects differently the μ-H and t-H isomeric forms. Thus, μ-H of highly substituted derivatives with bulky PR_3 in their oxidated $[Fe^{II}Fe^{II}]$ state is unstable toward reduction, compared to analogues with R of smaller size. The intramolecular strain thus generated can be released upon one-electron reduction in two ways depending on which one of the two rotational stereoisomers of μ-H is being reduced. The main point is that only μ-H shows release of the intramolecular strain because in t-H, a [Fe]-[Fe] bridging CO, which is strong π-acceptor and *trans*-oriented to two strong σ-donors, forms two strong [Fe]−C(O)−[Fe] bonds, preventing the two $[Fe]-L_3$ moieties from moving away from each other. Therefore, t-H of electron-rich species simultaneously bearing bulky PR_3 ligands cannot release the intramolecular strain, whereas μ-H does. This justifies the inversion of the $\Delta E°_{tH-\mu H}$ sign observed in **VI**-like systems. In the case of neutrally charged σ-donors, less strong than the apical CH_3S^- (such as NHC), our results again show a $\Delta E°_{tH-\mu H} < 0$, although t-H's of such derivatives show structural parameters indicating either $[Fe]-(P)R_3$ or [Fe]−[Fe]/[Fe]-H bond breaking.

Justification of such a result requires therefore generalization of the model based on the strain presence/absence and on how/if it can be released upon reduction. The general factor underlying the positivity or negativity of $\Delta E°_{tH-\mu H}$ is the similarity or diversity of the reductive behavior by t-H versus μ-H of a given species. Therefore, in the case of species with NHC and PMe_3, their t-H is strain-free already at the oxidized state because two such ligands are not able to make sufficiently strong [Fe]−(C)O bonds, which would preserve the unstable configuration. By contrast, μ-H of the same species is affected by intramolecular repulsion at the oxidized state, reduction of which allows alleviation. This implies that the behavior of t-H versus μ-H is once again different, although according to another pattern compared to that emerged in CH_3S^--based species. It is worth pointing out that the possibility of replacing the anionic CH_3S^- with neutrally charged σ-donors is a relevant factor because it favors the lowering of the reduction overpotentials.

Reducing the overall charge of the diiron dithiolates by, for instance, replacing neutral pdt with nitrogen-containing linkers that are susceptible to being protonated by acids, tends to favor t-H reduction over μ-H ($\Delta E°_{tH-\mu H}$ is shifted toward more positive values). The same effect is observed when calculating $\Delta E^{\circ}_{tH-\mu H}$ after protonation of the apical CH_3S^- in models of the diiron core of the enzyme active site. These observations, along with the necessity of incorporating a large number of donor ligands in the diiron dithiolate, before the repulsive strain effect pops up, implies that non negligible overpotentials are to be taken into consideration. Therefore, a compromise that can balance the *overpotential versus* ΔE° competition has to be pursued.

2.1.3 Key electronic features for H_2 binding in [FeFe]-hydrogenases

2.1.3.1 Introduction

Considering the catalytic cycle of the [FeFe]-hydrogenases in the direction of H_2 oxidation, the redox state directly involved in H_2 binding is referred to as H_{ox}, where the redox state of the H-cluster can be formally described as $[2Fe^{II}2Fe^{III}]$-$[Fe^{II}Fe^{I}]$. In addition, in H_{ox} a CO ligand bridges the two iron ions, and the $[Fe_d]$ atom is five-coordinated.[40, 41, 21, 42, 43] Due to the remarkable catalytic activity of [FeFe]-hydrogenases, the H_{ox} state has become a reference point for the design of biomimetic catalysts for H_2 oxidation. In fact, a variety of $[Fe^{II}Fe^{I}]$ complexes were synthesized and characterized in recent years.[98] However, despite their similarity to the $[2Fe]_H$ subcluster, $[Fe^{II}Fe^{I}]$ complexes are unable to catalyze H_2 oxidation,[98] unless suitable redox-active ligands mimicking the functional role of the [4Fe4S] cluster are covalently linked to the diiron moiety.[80] In any case, the catalytic activity of such biomimetic complexes is much lower than that of the natural enzyme. General considerations about catalysis [99, 100] suggest that in an efficient catalyst for H_2 oxidation the energy difference between intermediate species should be small, because the formation of very stable or very unstable transient species is not compatible with a fast catalytic process. For analogous reasons, the oxidation potential of the catalyst-H_2 adduct and of the other intermediate species formed in the catalytic cycle should be as close as possible to the corresponding potential for the uncatalyzed reaction. In addition, the investigation of both [FeFe]-hydrogenases and synthetic complexes has revealed that the presence of a suitable pendant amine group, which can play the role of a base in the heterolytic cleavage of H_2, greatly increases catalytic activity.[97] With the aim of

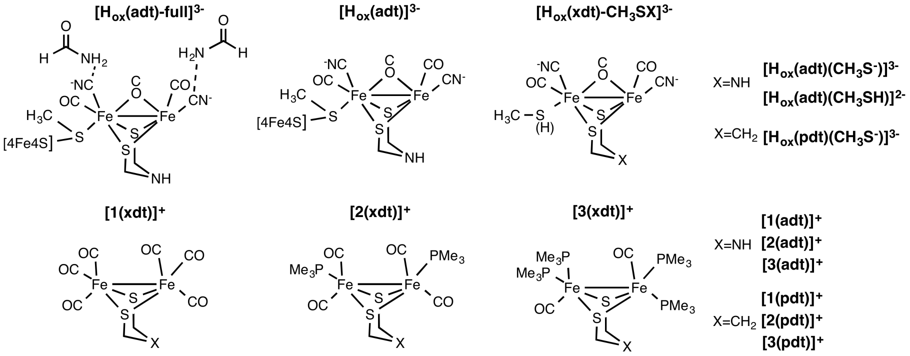


Figure 2.1.12: Schematic structure of the [FeFe]-hydrogenase $[Fe^{II}Fe^{I}]$ biomimetic models that have been investigated.

contributing to the design of novel and more efficient biomimetic catalysts for H_2 oxidation, we have used DFT to quantitatively compare key reaction steps in the catalytic cycle of [FeFe]-hydrogenases and related biomimetic complexes: H_2 binding and oxidation.

2.1.3.2 Methods

All DFT calculations have been carried out with the TURBOMOLE suite of programs,[57] applying the resolution of identity technique,[58] using the all-electron TZVP basis set [59] and the BP86 pure functional.[60, 61] This level of theory has been shown to be suitable and reliable for the investigation of [FeFe]-hydrogenase models.[62, 63, 48, 56, 64] Calculations were also performed with the hybrid B3LYP functional to validate the results obtained.[101, 102] The solvents (protein environment, water and CH_2Cl_2) polarizing effect have been modeled according to the COSMO approach,[85, 86] by considering a polarizable continuum medium characterized respectively by a $\varepsilon = 4.80$ and 9.01.

2.1.3.3 Result and discussion

Initially, the binding of H_2 molecule has been studied on a series of models of the H-cluster in the H_{ox} redox state. $\mathbf{[H_{ox}(adt)\text{-}full]^{3-}}$ model differs from the H-cluster only because the cysteine side chain bridging the [4Fe4S] and the $[2Fe]_H$

	ΔE		d [H]-[H]	d N/C(H)-(H)H
compound	ε=1	ε=4		
[H$_{ox}$(adt)-full]$^{3-}$	-2.26	-1.70	0.825	2.13
[H$_{ox}$(adt)]$^{3-}$	0.87	0.89	0.822	2.13
[H$_{ox}$(adt)(CH$_3$S$^-$)]$^{2-}$	1.46	0.42	0.816	2.19
[H$_{ox}$(adt)(CH$_3$SH)]$^{2-}$	3.46	2.69	0.808	2.28
[H$_{ox}$(pdt)(CH$_3$SH)]$^{2-}$	6.05	5.13	0.815	2.89

	ε=1	ε=9.01		
[1(adt)]$^+$	-3.24	-4.73	0.839	1.84
[2(adt)]$^+$	0.95	1.66	0.819	1.98
[3(adt)]$^+$	6.16	6.54	0.814	2.05
[1(pdt)]$^+$	2.02	2.83	0.810	2.92
[2(pdt)]$^+$	4.87	4.65	0.797	2.82
[3(pdt)]$^+$	7.50	6.98	0.800	3.08

Table 2.11: H_2 binding energies (in kcal mol^{-1}) computed with the BP86 and B3LYP functionals, and selected geometry parameters (from BP86 calculations) for models of the [FeFe]-hydrogenase active site and for [Fe^{II}Fe^{I}] biomimetic models.

subclusters is replaced by a methylthiolate group. Since in the enzyme the two CN$^-$ groups of the H-cluster are involved in hydrogen bonding interaction with the peptide chain, two formamide molecules forming hydrogen bonds with the CN$^-$ ligands have been also included. The simpler model **[H$_{ox}$(adt)]$^{3-}$** differs from the previous one for the absence of the formamides molecules. In **[H$_{ox}$(xdt)(CH$_3$X)]$^{2-}$** complexes the nature of the xdt strap has been varied betreen adt and pdt; in addition different protonation states of the methil thiol group have been testes. Three additional complexes with an increased basicity of the ligands at the diiron core and a different nature of the xdt strap have been studied [from **[1(xdt)]$^+$**to **[3(xdt)]$^+$]**, see Figure 2.1.12).

In all models, H_2 binds in a non-classical way, and the [H]-[H] bond elongation is generally proportional to the binding affinity (Table 2.11).

The computed binding energy of H_2 to the [Fe_d] atom of the **[H_{ox}(adt)-full]**$^{3-}$ model is -2.26 kcal mol^{-1} (considering a dielectric constant $\varepsilon = 1$), whereas H_2 binding to the simplified models **[H_{ox}(adt)]**$^{3-}$-, **[H_{ox}(adt)(CH_3S^-)]**$^{3-}$ and **[H_{ox}(xdt)(CH_3SH)]**$^{2-}$ is endothermic. The comparison of H_2 binding energies in **[H_{ox}(adt)(CH_3SH)]**$^{-}$ and **[H_{ox}(pdt)(CH_3SH)]**$^{-}$ reveals that the presence of the amino group in the chelating dithiolate ligand favorably contributes to H_2 binding (by about 2.6 kcal mol^{-1}; Table 2.11). Changing the dielectric constant ε from 1 to 4 (a value commonly used to model protein environment) generally leads to slightly more favorable binding energies (by about 1 kcal mol^{-1}; Table 2.11). The larger affinity of H_2 for **[H_{ox}(adt)-full]**$^{3-}$ and **[H_{ox}(adt)]**$^{3-}$, in comparison to diiron models of the active site, reveals that the presence of the [4Fe4S] cluster affects substrate binding. In fact, while binding of H_2 to the investigated models is always concomitant with delocalization of the unpaired electron from [Fe_d] to [Fe_p], in **[H_{ox}(adt)-full]**$^{3-}$ and **[H_{ox}(adt)]**$^{3-}$ partial electron transfer from the [2Fe]$_H$ to the [4Fe4S] cluster (about 0.13 charge units) is also observable. In addition, the binding energy difference between the two species reveals that also the H-cluster environment can affect substrate binding and in particular the formation of H-bonds between the CN^- ligands and the environment decreases the electron density on the [2Fe]$_H$ subcluster, favoring H_2 binding. Binding energies for a series of biomimetic [$Fe^{II}Fe^{I}$] complexes which differ for the characteristics of the ligands coordinated to each Fe atom and for the nature of the xdt strap were then computed. Computed H_2 binding energies for the symmetric diiron complexes (Table 2.11) range from exothermic (–3.24 kcal mol^{-1}; complex **[1(adt)]**$^{+}$) to strongly endothermic (+7.50 kcal mol^{-1}; complex **[3(pdt)]**$^{+}$), and are inversely correlated to the electron donor character of the ligands. More in detail, the more electron donor are the ligands, the lower is the affinity for H_2. The comparative analysis reveals that also in biomimetic complexes the nitrogen group of adt contributes to H_2 binding, as deduced by the comparison of binding energies in corresponding adt- and pdt-containing complexes (Table 2.11). In addition, both in active site and biomimetic models, the [H]-[H] and N(H)-(H)H distances are often inversely correlated (Table 2.11), indicating that the interaction between the N atom of adt and one of the two H atom of the coordinated H_2 molecule results in activation of the latter.

In relation to the catalytic properties, it should be noted that a too high affinity of the catalyst for H_2 would be detrimental, because the formation of very stable intermediate species necessarily implies large energy barriers in the next catalytic steps. On the other hand, a too low affinity for the substrate, as computed in **[3(pdt)]**$^+$) implies that the catalytic process could take place efficiently only under high H_2 pressure. In line with these remarks, in the most complete H-cluster model **[H$_{ox}$(adt)-full]**$^{3-}$ the computed H_2 binding reaction energy is close to 0 kcal mol^{-1}. When considering biomimetic complexes, reaction energies close to 0 kcal mol^{-1} are obtained to a lesser extent for complex **[2(adt)]**$^+$. When comparing the enzyme cofactor and biomimetic models, it is also remarkable that the highly negative overall charge of the H-cluster (-3), compared to biomimetic complexes (+1), does not impair H_2 binding. The apparently contradictory observation that H_2 can bind via σ-donation to a highly anionic complex, can be rationalized considering that only the five-coordinated iron atom of the metal cofactor must be sufficiently electron poor to allow H_2 binding. In fact, a relatively electron-poor five-coordinated Fe ion can be present also in highly anionic species (such as the H-cluster) when suitable electron acceptor ligands (such as CO and CN-) are coordinated to the Fe ions, thus draining electron density away from metals.

2.1.3.4 Conclusion

The catalytic cofactor of [FeFe]-hydrogenases, the so-called H-cluster, has inspired the synthesis of hundreds of biomimetic complexes. In fact, the elucidation of structure-activity relationships in this class of iron coordination compounds is expected to drive the design of affordable catalysts for H^+ reduction and H_2 oxidation. Here the fundamental factors for H_2 efficient binding have been analyzed and discussed. Our analysis highlights that a sufficiently electron poor diiron subcluster is able to allow H_2 coordination. In addition, the amine group contained in the adt strap between the two iron atoms is crucial not only for its well documented role as a base in H_2 heterolytic cleavage,[103] but also for tuning the H_2 affinity to the diiron cluster.

2.2 Synthetic catalysts for hydrogen storage and carbon dioxide reduction

The efficient and catalytic reduction of CO_2 by natural photosynthesis has inspired many sustainable energy and clean tech applications.[104, 105, 106] In addition CO_2 is also a greenhouse gas and its fixation as part of a carbon-neutral energy cycle would significantly decrease its associated environmental risk.[107] As explained in the previous section, H_2 is desirable as a fuel because it can be converted efficiently to energy without producing toxic products or greenhous gases, and the rates and efficiencies with which it can be obtained from water with sunlight and the efforts made by the scientific community to find eco-friendly way to produce it continue to increase.[108, 109] Unfortunately, on an industrial scale, H_2 is constrained by its physical properties, leading to safety concerns, transport problems, and low energy density.[104] Many different approaches based on H_2 storage devices like cryogenic or high density gas containers have been tested, but the results have not proved economically viable.[110] In contrast, liquids with a high hydrogen content ($> 4\%$) can be safe to handle and offer a greater energy density. In addition, their transport can be feasible with the existing infrastructure for gasoline and oil. In this context, aqueous formic acid (HCOOH) and its conjugate base formate ($HCOO^-$) are promising liquid hydrogen-storage media because of the low energy barrier to their formation starting from H_2 and CO_2 (ΔG_{298K}=0.96 kcal mol^{-1}).[111, 112, 113] The equilibrium given by the reaction

$$H_2\ (aq.) + CO_2\ (aq.) \rightleftarrows HCOOH\ (aq.) \qquad (2.2.1)$$

can be interpreted as a hydrogen storage system, with fuel ready H_2 on the left part and stored H_2 on the right. It has to be underlined that the value added chemical HCOOH has some intrinsic properties that can be exploited in the industrial sector. First of all formic acid itself is an important preservative for winter feed for cattle and an antibacterial product in the poultry sector.[114] Furthermore, HCOOH is commonly used as a C1 feedstock for the production of solvents and synthetic building blocks, but this topic will not be presented in this manuscript.[115]

One way to catalytically produce formic acid starting from H_2 and CO_2 is the use of biological systems and in particular isolated enzymes as formate dehydrogenase (or FDH). FDH from *Candida boidinii* is extremely interesting from an industrial point of view for its capability of regenerating NADH from NAD^+ and formic acid, yielding CO_2 as a by-product. Unfortunately this FDH discovered by the pioneering work of Kula and colleagues in 1976 [116] was found to be almost inactive in the reverse reaction, required to make formic acid from CO_2. More recently alternative FDHs have been described and in particular FDH of *Clostridium carboxidovorans* has been found to be oxygen-tolerant and higly active in the NADH oxidation.[117] Furthermore, this enzyme was found to have lower affinity for formate, making it more useful for the production of this latter. Another enzyme found in *Clostridium autoethanogenum* is a NADH-dependent FDH, part of a complex composed of seven subunits.[118, 119] One of the other subunits is an [FeFe]-hydrogenase and together all the subunits catalyze the reversible coupled reduction of ferrodoxin and $NADP^+$ with H_2 or formate and the reversible formation of H_2 and CO_2 from formate.[118, 119] The major disadvantage is that all these FDHs require NAD(P)H in stoichiometric amounts and consequently a cost increase if scaled-up to industrial level. Recently other intresting enzymes like hydrogen-dependent carbon dioxide reductase (HDCR) have been discovered, but their oxygen-sensitivity puts a huge challenge on the development of an efficient whole-cell system under anaerobic condition.[120]

A practical and promising solution to overcome these problems, is the development of synthetic functional analogues of FDH enzyme, able to store CO_2 and H_2 in formic acid or formate. The first report of CO_2 hydrogenation to formic acid dates back to 1935, when Farlow and Atkins used Ni-Raney as a catalyst.[121] In 1976,[111] the first homogenous catalyst for CO_2 reduction and formate generation was described, and since then several papers were published on this trend topic.[112, 122, 105] Looking at the equation 2.2.2

$$CO_2\ (aq.) + e^- \rightleftarrows CO_2^{-\bullet}\ (aq.) \qquad E°\text{=-1.9 V} \tag{2.2.2}$$

it is clear that the one electron reduction of CO_2 requires very negative potentials, due to the energy required for structural rearrangement of the linear CO_2 molecule to form the bent CO anion radical.[123] The high overpotential associated to this reaction remains the major obstacle to rapid and electrochemical reduction of CO_2.[124] The optimal catalyst for CO_2 reduction should avoid the highly energetically unfavourable formation of an anion radical as a catalytic intermediate, favoring lower energy processes for formic acid production. Recent promising results have been achieved with the development of iridium,[113, 125] rhodium, [126, 127] and ruthenium catalysts.[112, 128, 129] Of course, the use of homogenous catalysts containing precius metals like those mentioned before is not economically affordable particularly at industrial level. Nowadays, only few non-noble metal-based systems are known for the selective reduction of CO_2. The use of biorelevant metals such as iron, cobalt or nickel has in fact been scarcely investigated compared to noble-metal-based complexes. Even if recently improvements were reported by the work of Beller [130] and Milstein,[131] all the non-noble metal catalysts show still low activity.[132, 133, 134]

Since the use of cheap and abundant metals in synthetic chemistry is an essential requirement for the design of homogenous catalysts able to store H_2 and CO_2 in molecules of formic acid or formate in a eco-friendly way, in next section a DFT analysis on reactivity of the most efficient iron based catalyst for CO_2 hydrogenation developed by Beller and colleagues [135] will be presented. The aim is to understand which molecular determinants underlies the observed reactivity.

2.2.1 On the reactivity of the most efficient iron-based catalyst for CO_2 hydrogenation

2.2.1.1 Introduction

As mentioned before, iron-based electron reduction of carbonate and carbon dioxide constitutes a challenging and highly attractive goal. Beller and co-workers described for the first time in 2010 [114] an active iron catalyst system which can be used for the reduction of both carbon dioxide and bicarbonate to give formate and derivatives. The success of this pioneering work comes from the combination of the iron source $Fe(BF_4)_2 \cdot 6H_2O$ and the tridentate ligand $P(CH_2CH_2PPh_2)_3$ or $(P_{Et}P_3)$ to form the iron-hydride complex $[FeH(P_{Et}P_3)]BF_4$. Here the CO_2 hydrogenation proceed with an excellent yield of 88% and an high turnover number (TON, 610 moles of sodium formate per mole of catalyst). The same complex has been tested in 2011 in the direction of formic acid dehydrogenation.[136] Presynthesized hydride catalyst $[FeH(P_{Et}P_3)]BF_4$ showed an enhanced catalytic activity (TON=1923) com-

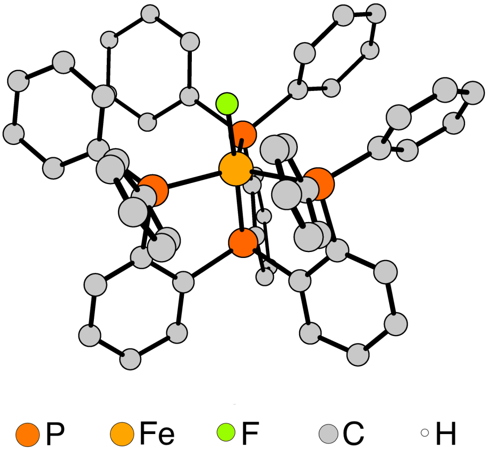


Figure 2.2.1: Molecular structure of complex [FeF(PP$_3$)]BF$_4$. For clarity hydrogen atoms and BF$_4$ counteranions have been omitted.

pared to the *in situ* assembled system. The most efficient iron-based catalyst for CO$_2$ hydrogenation was presented in 2012.[135] Instead of using a tridentate ligands, the iron atom was this time coordinated by a multidentate phosphorous chelate. Specifically the idea was to use a tris(2-diarylphosphino)aryl)phosphines ligand or (P$_{Ph}$P$_3$). The reaction of the iron source Fe(BF$_4$)$_2$.6H$_2$O with the phosphorous ligand generates a [FeF(P$_{Ph}$P$_3$)]BF$_4$ (Figure 2.2.1). The hydrogenation of CO$_2$ in methanol (MeOH) was investigated in presence of amines. The direct hydrogenation of CO$_2$ in presence of triethylamine led to a mixture of formic acid and methyl formate with a conversion of CO$_2$ of around 15% and a TON of 1692 moles converted per moles of catalyst. To understand the mechanism of the Fe-catalyzed hydrogenation nuclear magnetic resonance (NMR) measurements of different catalytic experiments were performed. Figure 2.2.2 illustrates the catalytic mechanism for iron-catalyzed hydrogenation of CO$_2$.

The initial formation of a hydrido di-hydrogen complex **B** is proposed. This is the result of the hydrogenation of the original complex [FeF(P$_{Ph}$P$_3$)] at 80°C within a few hours and its behaviour is similar to the one of the known analogue with (P$_{Et}$P$_3$) ligand.[130] When CO$_2$ is added to the reaction solution a second NMR hydride signal appears, even if no formation of formic acid or formate is observed. This observation can be explained through the formation

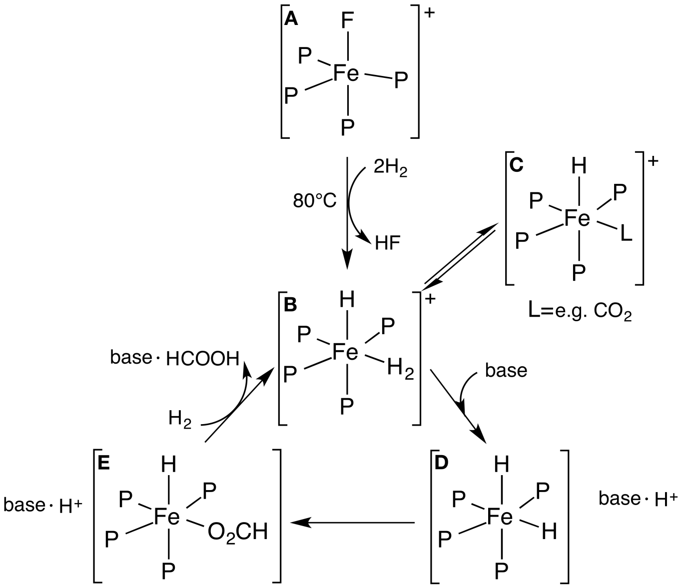


Figure 2.2.2: Postulated mechanism for the iron-catalyzed hydrogenation of CO_2.

of an equilibrium between complexes **B** and **C**, in which the hydrogen ligand is exchanged either with solvent molecule or carbon dioxide. Only after the addition of the base triethylamine a product signal appears with the disappearance of complex **B**. It has been postulated that the presence of the base is necessary to convert **B**, a resting state of the catalytic species, into the reductively active form of the catalyst, **D**. The insertion of CO_2 into the [Fe]-[H] bond forms the corresponding iron-formate complex **E**. The final protonation of this latter and the coordination of hydrogen closes the cycle and formic acid can be detected by NMR spectroscopy. **C** has been postulated to be the starting point for a second competitive cycle in the analogous system $[FeH(P_{Et}P_3)]BF_4$. Here instead, this intermediate is described as dead-end intermediate, without any catalytic activity.

Since a detailed mechanism for $[FeF(P_{Ph}P_3)]BF_4$ catalytic CO_2 hydrogenation is still missing, here it is presented a DFT investigation of all the alternative reaction pathways. The solvent presence on the energy profile has also been taken into account, as well as the importance of the base and also a possible role of the counterion. This study allows to rationalize the different mechanisms proposed for [FeF($P_{Ph}P_3$)]BF_4 and [FeH($P_{Et}P_3$)]BF_4, in order to justify the presence of two competing cycles in the latter system that can't be detected in the former.

2.2.1.2 Methods

All DFT calculations have been carried out with the TURBOMOLE suite of programs,[57] applying the resolution of identity technique,[58] using the all-electron TZVP basis set [59] and the BP86 pure functional.[60, 61] Atomic coordinates of the starting [Fe(II)-fluoro-tris(2(diphenylphosphino)phenyl) phosphino tetrafluoroborate] structure were obtained from X-ray structure.[135] In order to enclude entropy, zero-point energy (ZPE) and enthalpy contribution into the self-consisted field energy (E_{SCF}), free energy values (G) were calculated considering three different contributions to the total partition function ($q_{translational}$, $q_{rotational}$ and $q_{vibrational}$).[19] The T and P values were generally set at 353.15 K and 60 bar respectively to reproduce reaction conditions. Because of the high values of the SCF wavenumbers arising from the harmonic approximation, a scaling factor of 0.9914 was applied.[57] Solvent (MeOH) has been modeled according to the COSMO approach,[85, 86] by considering a polarizable continuum medium characterized by ε value of 33. Transition states search has been performed according to pseudo Newton-Raphson procedure. Initially, geometry optimization of a guessed transition state structure is performed constraining the distances corresponding to the reaction coordinates. A vibrational analysis is then carried out on the optimized structure: the presence of a negative eigenmode corresponding to the reaction coordinate confirms the nature of the transition state. 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, the first one being that associated to 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 at the previous step. In light of available experimental data and considering the chemical nature of the ligands of the iron atom, both low-spin and high-spin species have been taken into account.

The evaluation of the efficiency of the catalyst (the turnover frequency or TOF) was performed applying the innovative approach proposed by Kozuch and Shaik.[99, 100] With this method, the TOF of a cycle does not depend on the rate determining step but it is rather an integrated rate function of the entire cycle (see equation 2.2.3)

$$\mathrm{TOF} = \frac{k_{\mathrm{B}}\mathrm{T}}{h}\mathrm{e}^{\delta E/RT} \tag{2.2.3}$$

The apparent activation energy of the whole cycle, the so-called energetic span (δE), is influenced by the energy levels of two species: the TOF determining transition state (TDTS) and the TOF determining intermediate (TDI) (see equation 2.2.4).

$$\delta E = \begin{cases} \mathrm{TDTS} - \mathrm{TDI} & \text{if TDTS appears before TDI} \\ \mathrm{TDTS} - \mathrm{TDI} + \Delta \mathrm{G_r} & \text{if TDTS appears before TDI} \end{cases} \tag{2.2.4}$$

where $\Delta \mathrm{G_r}$is the reaction energy.

TDTS and TDI are not necessarily the highest and the lower state respectively of the catalytic cycle, but are the two species that maximize the energetic span and thereby gauge the kinetics of the catalyst.[99, 100]

2.2.1.3 Results

The main aim of the present work is the dissection of the two plausible reaction mechanisms for CO_2 hydrogenation catalyzed by the first cheap and efficient iron-based $[FeF(P_{Ph}P_3)]BF_4$ system.[135] NMR studies [135] have revealed that **1**, the starting intermediate of the cycle, is in equilibrium with another species in which dihydrogen is exchanged either with a solvent molecule or carbon dioxide **7** (see Figure 2.2.3; the same species in Figure 2.2.2 are named **B** and **C**). Species **1** is shown to be unreactive before the addition of the base, a step necessary to convert **1** to the reactively active species. Since **7** is the starting point of cycle 2 proposed by Beller's group for the analogous complex $[FeH(P_{Et}P_3)]BF_4$,[114, 136] we decided to take into account both catalytic cycles in order to investigate all the alternative hydrogenation mechanisms for $[FeF(P_{Ph}P_3)]BF_4$ and disclose the reason why the catalyst can not proceed via cycle 2 reactions.

All intermediates and transition states (TS) found are cationic species and the

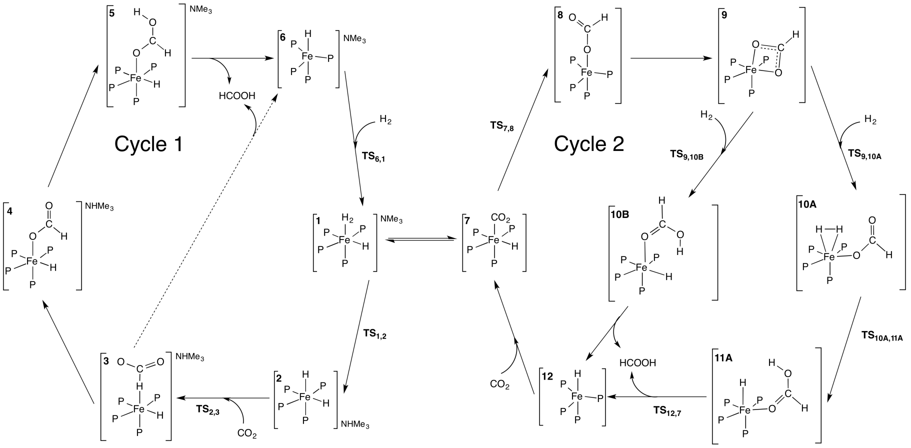


Figure 2.2.3: Proposed mechanism for [FeH($P_{Ph}P_3$)]BF_4mediated CO_2 hydrogenation.

presence of the base NEt_3 (here modeled as NMe_3 to reduce the computational cost), fundamental for the catalysis, is explicitly considered only in cycle 1. The effect of the counterion and of the base on the catalytic activity will be discussed later. Reaction pathways are depicted in Figure 2.2.4

Cycle 1 starts with the deprotonation of the hydrido-dihydrogen complex **1** to give the dihydride complex **2** and $NHMe_3^+$. The computed free energy barrier for deprotonation of the classical hydride is about 15 kcal mol^{-1} higher than the one of the more acidic and less hindered dihydrogen ($\mathbf{TS}_{1,2}$ $\Delta G^{\neq} =$ 14.71 kcal mol^{-1}, Figure 2.2.5). Catalysis proceeds through the attack of a CO_2 molecule to one of the *cis*-hydride ligand of **2**, through a *direct hydride-transfer* transition state $\mathbf{TS}_{2,3}$ (Figure 2.2.5) with a barrier of 14.90 kcal mol^{-1}. In **3** the formate is η^1-coordinated to Fe through a weak [Fe]-[H] interaction (the Fe-H distance is 1.72 Å) which is mostly stabilized by the presence of the protonated base. Even if formate dissociation from adduct **3** is only 3.83 kcal mol^{-1} energetically uphill, the formation of more stable η^1-O isomers **4** is feasible. At this point the release of HCOOH after protonation (from **5**) becomes less favorable. $\mathbf{TS}_{6,1}$ is only 1.91 kcal mol^{-1} higher than **6**, suggesting a fast H_2 binding to form the stable hydrido-dihydrogen species that closes the catalytic cycle.

Cycle 2 illustrates the formation of formic acid through CO_2 hydrogenation by a so-called *migratory-insertion-like* mechanism, already proposed for the

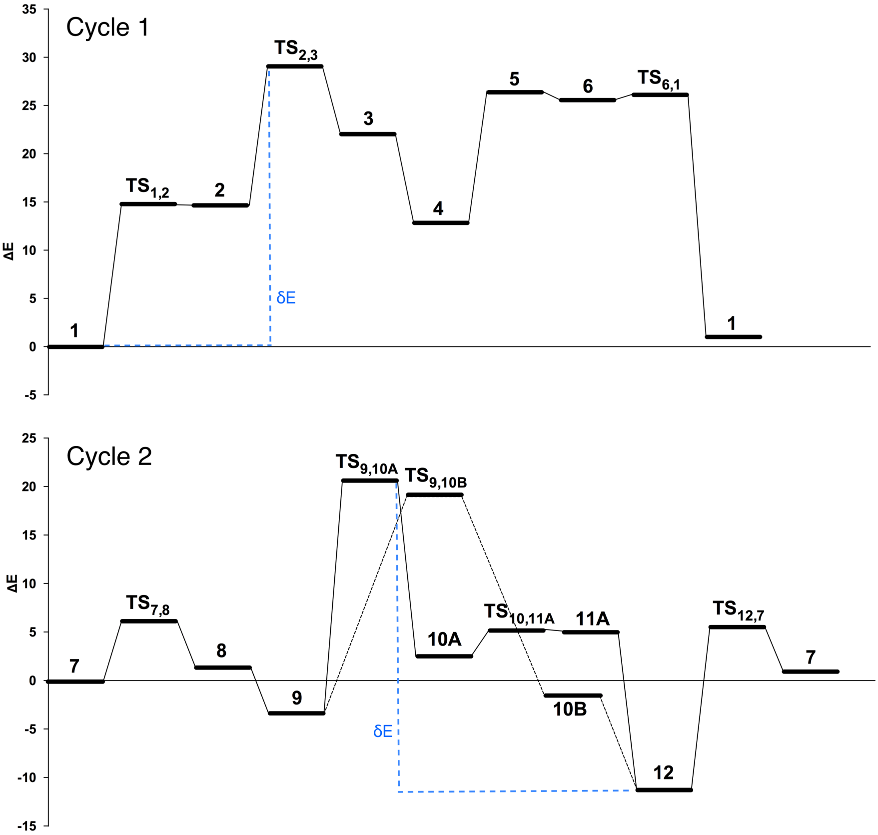
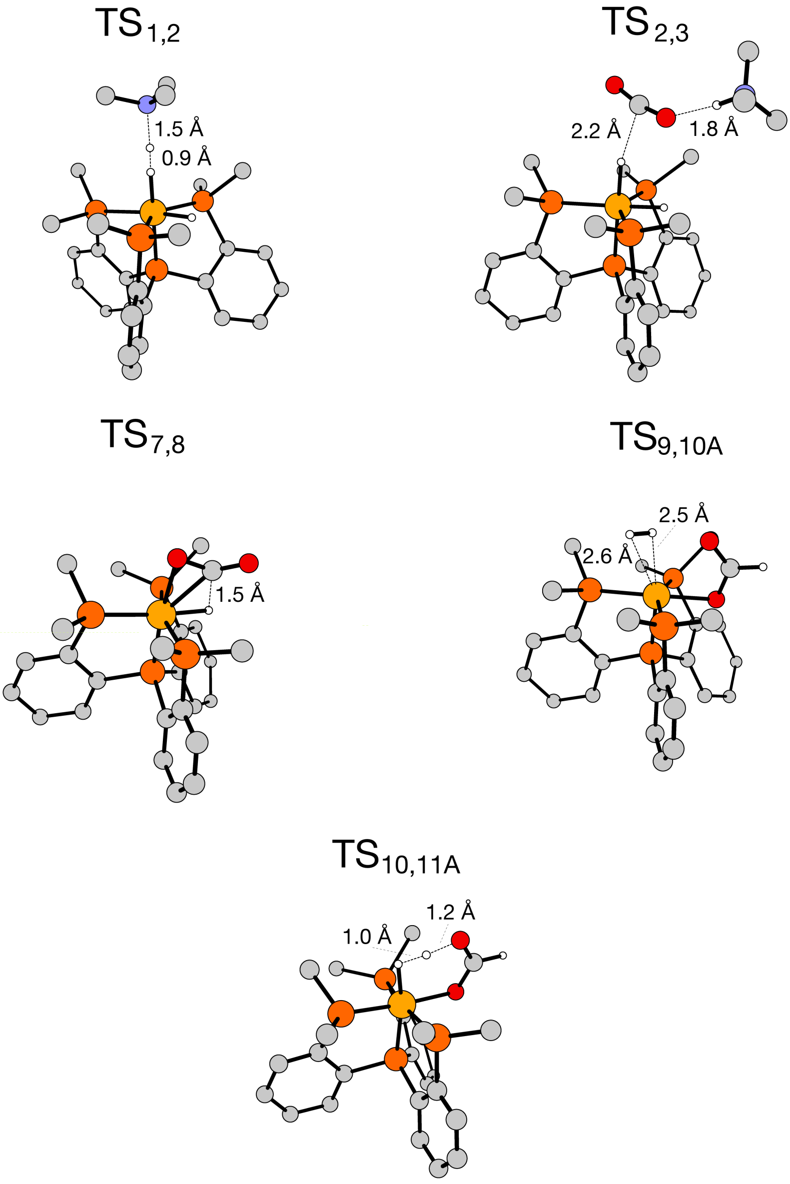


Figure 2.2.4: Free energy profile of cycle 1 and cycle 2.



$TS_{1,2}$ $TS_{2,3}$ $TS_{7,8}$ $TS_{9,10A}$ $TS_{10,11A}$

Figure 2.2.5: Some of the most interesting transition states of both cycle 1 and 2.

$[FeH(P_{Et}P_3)]BF_4$ complex developed by Beller and coworkers,[114, 136] without the assistance provided by the base NMe_3. The formation of **7**, starting from **1** was suggested by NMR data. Among all possible binding modes of CO_2 to the iron atom, the η^1-O coordination (Figure 2.2.3) has been found to be the most stable. The *migratory-insertion-like* mechanism, which takes place through the binding of the hydride to the carbon atom of CO_2 in $\mathbf{TS}_{7,8}$, features a barrier of 6.20 kcal mol^{-1}. In $\mathbf{TS}_{7,8}$ [C]-[H] and [Fe]-[O] distance are 1.540 Å and 2.074 Å respectively (Figure 2.2.5). The binding of formate to the iron atom yielding **8** is stable and the penta-coordinated Fe^{II} is characterized by a triplet spin state (S=1). The isomerization of **8** to **9** is associated by a downhill step of 4.84 kcal mol^{-1}. **9** is characterized by a η^2-O coordination to the Fe atom through two [Fe]-[O] bonds of 2.090 and 2.058 Å. From **9** two different reaction pathways are possible depending on the binding position of a H_2 molecule (see Figure 2.2.3). $\mathbf{TS}_{9,10A}$ (Figure 2.2.5) shows the binding of H_2 to the Fe atom in the apical position of $[FeH(P_{Ph}P_3)]$ ($\Delta G^{\neq} = 24.02$ kcal mol^{-1}) while $\mathbf{TS}_{9,10B}$ displays a side-on binding mode ($\Delta G^{\neq} = 22.41$ kcal mol^{-1}). Notably, these two similar transition states evolve to completely different products. While intermediate **10A** shows both the formate anion and the dihydrogen molecule bound to the metallic center, **10B** features a formic acid molecule bound to the iron-hydrido center (Figure 2.2.3). From **10A** the formation of the formic acid takes place through the deprotonation of the H_2 molecule in $\mathbf{TS}_{10,11A}$ (Figure 2.2.5), higher in energy by 2.79 kcal mol^{-1}. In **11A** the HCOOH is η^1-coordinated to the Fe-center (Figure 2.2.3). From both **11A** and **10B** the release of the formic acid takes place spontaneously to regenerate a $[FeH(P_{Ph}P_3)]$ complex (**12**, $\Delta G^{\neq} = -16.20$ and $\Delta G^{\neq} = -9.54$ kcal mol^{-1} respectively). Starting again from the iron-hydrido complex, the binding of a single CO_2 molecule takes place through $\mathbf{TS}_{12,7}$ ($\Delta G^{\neq} = 16.59$ kcal mol^{-1}), characterized by a [Fe]-[O] distance of 3.602 Å.

Free energy profiles of cycle 1 and cycle 2 show that the energetic spans (δE) are 29.20 kcal mol^{-1} ($\mathbf{1} \rightarrow \mathbf{TS}_{2,3}$) and 30.50 kcal mol^{-1} ($\mathbf{12} \rightarrow \mathbf{TS}_{9,10B}$) respectively. According to DFT results, catalysis does not seem to follow a preferential pathway (δE value is the same in both cases), suggesting that a purely kinetic analysis is not sufficient to fully rationalize experimental observations concerning the catalytic mechanism. However, the competition between the two proposed mechanisms can be also explained in terms of $[FeH(P_{Ph}P_3)]$ affinity for H_2 versus CO_2, therefore by using thermodynamic criteria. It was expected that the more the H_2 exchange with a CO_2 molecule is energetically unfavored, the more cycle 2 will be disfavoured compared to cycle 1. Thus, ΔG value associated with the reaction: $[FeH(H_2)(P_{Ph}P_3)]$ (**1**) + $CO_2 \leftrightarrow [FeH(CO_2)(P_{Ph}P_3)]$ (**7**) + H_2 were computed, and it was noticed that the equilibrium is completely shifted towards the reactants ($\Delta G = 11.2$ kcal mol^{-1}), preventing CO_2 binding. The equilibrium driving force is the great stability of *cis*-hydride-η^2-dihydrogen complexes with tetraphos ligand, due to the so called cis-effect, which has been widely unraveled by Bianchini's group [137, 138, 139] and also observed by Beller and coworkers.[114, 136] Since catalysis starts with the $[FeH(H_2)(P_{Ph}P_3)]$ intermediate, we can conclude that even if cycle 2 is kinetically equivalent to cycle 1, it is not thermodynamically accessible.

In order to make a direct comparison between the two iron-based Beller's catalysts,[114, 135] the aromatic tetraphos bridges ($P_{Ph}P_3$) was switched to the aliphatic ($P_{Et}P_3$) analogous and then δE values of cycle 1 and cycle 2 were computed. Also in the $[FeH(P_{Et}P_3)]$ case the two mechanisms are almost kinetically equivalent, since cycle 2 is preferred by only ca. 3 kcal mol^{-1} to cycle 1 (31.02 versus 34.53 kcal mol^{-1}). These results are in agreement with previous DFT studies,[136, 140] which highlighted that the HCOOH dehydrogenation reaction catalyzed by $[FeH(P_{Et}P_3)]$ is a competition between the two mechanisms which are very similar in energy.

To inspect in a more detail which determinants affect catalysis, we studied the effect of those species (base NEt_3 and counterion BF_4) that could interact with the catalyst or assist hydrogenation reaction steps. Given that TOF is governed by δE, we focused on the free energy changing of TDI and TDTS in both cycles. The presence of a base is necessary to enable cycle 1, as the direct-hydride transfer is possible only after deprotonation of **1**, which constitutes the most abundant intermediate of the cycle. In the deprotonation step the base is directly involved, and thus the free energy barrier will be lowered when using a stronger base; since NEt_3 has been modeled with the less basic NMe_3, the computed $\mathbf{TS}_{1,2}$ free energy (and consequently δE) could be slightly overestimated. In addition, the base is predicted to assist CO_2 binding, since $NHMe_3{}^+$ stabilize $\mathbf{TS}_{2,3}$ of about 2 kcal mol^{-1}. No minimum corresponding to a η^1-H coordination of formate has been observed, and the reaction evolves directly to a η^1-O species. In cycle 2 the base eases H_2 binding lowering the barrier of about 2 kcal mol^{-1}, avoiding the formation of intermediates **9** and directly producing intermediates **11A** ($\Delta G^{\neq} = 40.68$ kcal mol^{-1}).

Calculations to study the role of the counterion were performed on TDTS on both cycle 1 and 2, but no significant influence of BF_4 on catalysis has been observed.

2.2.1.4 Conclusion

DFT has allowed to determine free energy profiles associated with the intimate mechanism by which $[FeF(P_{Ph}P_3)]$ catalyzes CO_2-to-HCOOH reduction. This is crucial for the understanding of the molecular determinants characterizing the catalytic activity. The use of the energetic span parameter δE has revealed to be useful to analyze reaction kinetics, nonetheless inclusion of thermodinamic criteria is relevant to fully rationalize the experimental scenario. In order to understand the effect of the chelating ligand on stereo-electronic properties of the system and to predict if a further improvement on the activity of this promising catalyst is achievable, future studies on variation of the first and second coordination sphere of the metal will be performed.

Chapter 3

Computations meet experiments: spectroscopic studies of metal-bound transferrin.

> *All excess is ill, but drunkenness is of the worst sort.*
> *It spoils health, dismounts the mind, and unmans men. [...]*
> *In fine, he that is drunk is not a man:*
> *because he is so long void of Reason,*
> *that distinguishes a Man from a Beast.*
>
> *William Penn*

3.1 Physiology of transferrin

Iron is essential for all types of cells and organisms. On the other hand, too much iron is lethal. It is therefore necessary to precisely regulate its intracellular homeostasis and transport to diverse locations proteins and metalloenzymes. Transferrins are nonheme iron-transport glycoproteins, found in multicellular organisms from cockroaches to humans, that play the role of carrying iron ($\mathrm{Fe^{III}}$) from the sites of intake to the circulation system, the cells and the tissues. This family includes serum transferrin, ovotransferrin, found in egg white, and lactoferrin, present in the white blood cells and secretory fluids.[141]

Human serum transferrin (Tf) is composed of 679 amino acid residues and has a molecular weight of ∼79 kDa. The molecule is stabilized by 19 intra-chain

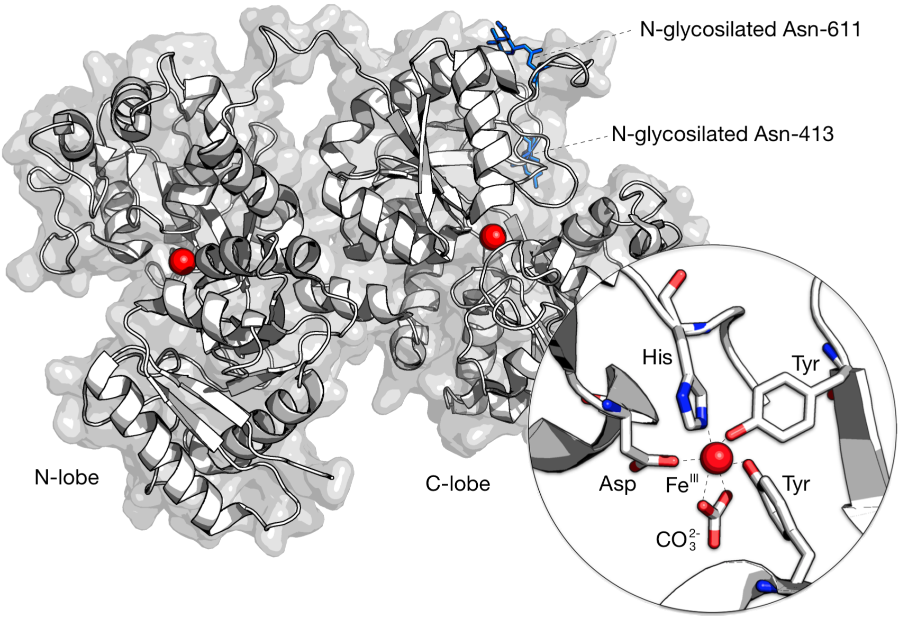


Figure 3.1.1: Tf structure and focus on the Fe^{III}binding site.

disulfide bonds and is protected by three carbohydrate side chains, two of which are N-linked (Asn-413 and Asn-611) and the third is O-linked (Ser-32). The Tf molecule is divided into two evolutionary related lobes, designated N-lobe (336 amino acids) and C-lobe (343 amino acids), which share a high degree of similarity (42%). Each lobe contains a metal binding site and is divided into two subdomains, CI and CII and NI and NII, respectively, where a central mixed β-sheet is surrouded by several α-helices. In each lobe, two subdomains are connected by a hinge consisting of two extended antiparallel β-strands, which form a cleft where the metal can be placed (see Figure 3.1.1).[142, 143]

The binding site in both lobes has four conserved amino acids including two tyrosines, one aspartic acid and one histidine (Asp-63, Tyr-95, Tyr-188, His-249 on the N-terminal lobe and Asp-392, Tyr-426, Tyr-517, His-585 on the C-terminal lobe). These residues are arranged in a distorted octahedral coordination. In addition, the binding site requires a bidentate ligand carrying two negatively charged oxygen atoms to further stabilize the metal complex. In serum Tf this essential role is played by a molecule of carbonate ion (CO_3^{2-}, Figure 3.1.1). The surrounding amino acid residues are thought to further stabilize the metal-binding site, and they have crucial roles in iron release. The structural characteristics of Tf binding site allow this protein to tightly but reversibly binds $\mathrm{Fe^{III}}$ (K_d of $\sim 10^{-22}$ M) and transports it from extracellular fluid to cytosol, thereby playing a critical role in the maintenance of human cellular $\mathrm{Fe^{III}}$ homeostasis as well as the prevention of damage from free radicals. Deficiencies in Tf lead to insufficient cellular $\mathrm{Fe^{III}}$, with a subsequent inhibition in cell growth and proliferation. It is noteworthy that in blood plasma only 30% of the Tf sites is saturated with $\mathrm{Fe^{III}}$. The remaining sites are available for other metals: therapeutic $\mathrm{Bi^{III}}$, radio diagnostic $\mathrm{Ga^{III}}$ and $\mathrm{In^{III}}$ and toxic $\mathrm{Al^{III}}$ and $\mathrm{Pu^{IV}}$ metals. These metal ions share similar coordination sites to $\mathrm{Fe^{III}}$ in Tf proteins, locating in the binding clefts formed within the two different lobes.[144]

Employing high-performance electrophoretic methods, like isoelectric focusing (IEF) followed by immunoblot and staining, Tfs solution is divided in a multitude of bands. This happens because Tf is not a group of homogenous molecules, but shows a distinct microheterogeneity due to the different iron load, the genetic variants and the glycan chians. From the analitical point of view this protein is a group of similar isoforms referred to as Tf.[141] As already mentioned, each Tf molecule can bind a maximum of two metal irons, preferably $\mathrm{Fe^{III}}$. Depending on the supply of the organism, Tf molecules are iron free (apo-Tf) or loaded with one ($\mathrm{Fe_{1N}}$- and $\mathrm{Fe_{1C}}$-Tf, where N and C indicate the N- and C-terminal lobes respectively) or two (holo-Tf) $\mathrm{Fe^{III}}$ ions. In healthy controls, Tf iron saturation is ∼30%, and apo-, $\mathrm{Fe_{1N/C}}$-, and holo-Tfs are detectable in serum. The isoelectric point (pI) of the Tf molecule decreases by ∼0.2 pH units with each $\mathrm{Fe^{III}}$ ion bound, permitting the separation of the distinct stoichiometric forms by electrophoresis.[141]

The two Tf N-glycan chains bound to Asn-413 and Asn-611 differ in their degree of branching, showing bi-, tri-, and tetra-antennary structures. Tf usually contains two complex carbohydrate chains consisting of four variable carbohydrates (N-acetylglucosamine, mannose, galactose, sialic acid). Sialic acid is the only charged carbohydrate, always positioned terminally.[145] The main isoform of Tf has a pI of 5.4 and four sialic acid residues (two on each carbohydrate chains) but minor isoforms with two (disialo-Tf, pI 5.7), three (trisialo-Tf, pI 5.6), five (pentasialo-Tf, pI 5.2) and six (hexasialo-Tf, pI 5.0) sialic residues have been identified in normal human serum. Also, traces of asialo-Tf and monosialo-Tf (pI 5.9 and 5.8, respectively) can be determined in serum (see

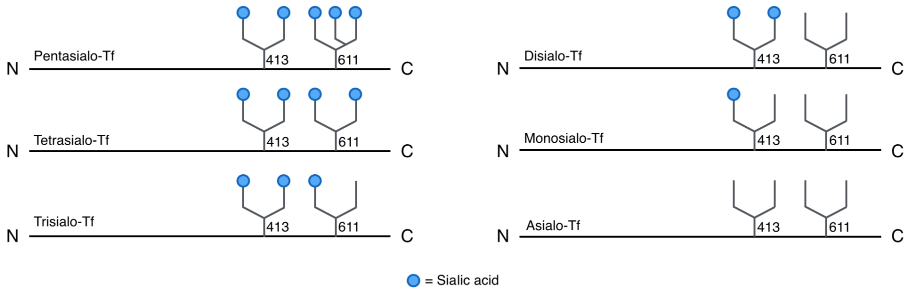


Figure 3.1.2: Microheterogeneity of Tf attributable to different N-glycan chains.

Figure 3.1.2).[146]

The relative amounts of these Tf isoforms as a percentage of total serum Tf are 1–3% hexasialo-Tf, 12–18% pentasialo-Tf, 64–80% tetrasialo-Tf, 4.5–9% trisialo-Tf, and $<$2.5% disialo-Tf in healthy persons. Asialo- and monosialo-Tf are not detectable or represent $<$0.5% (asialo-Tf) and $<$0.9% (monosialo-Tf), respectively, of total Tf under nonpathological conditions.[147] Carbohydrate-Deficient Tf (CDT) is the collective name of a group of minor isoforms of human Tf with a low degree of glycosylation.[148] The Tf glycoforms collectively referred to as CDT, on the basis of a large scientific consensus, include the isoforms with pI values ≥ 5.7, i.e. asialo-Tf, monosialo-Tf and disialo-Tf.[146]

3.1.1 Spectroscopic studies to improve CDT detection

3.1.1.1 Introduction

Since 1976, when Stibler and Kjellin reported an increased concentration of CDT isoforms in the cerebrospinal fluid and serum of alcoholics,[149] many studies have been published pointing out the high specificity and sensitivity of CDT for the diagnosis of chronic alcohol abuse. According to the current literature, it is widely accepted that CDT serum concentration increases after sustained alcohol intake (50–80 g/day) lasting for at least 7 days and decrease, after cessation of drinking, with a half-life of 14 days.[148] On this basis, CDT has been introduced as a diagnostic marker of chronic alcohol abuse and recently, according to a sound body of literature, it has proved to be the most reliable indicator, particularly in terms of specificity.[148]

Although an alcohol-induced increase of the less glycosilated isoforms of Tf has been observed already in 1976,[149] the exact pathomechanisms for the increase in CDT isoforms during chronic alcohol abuse are not completely understood yet. It is most likely that ethanol and/or its metabolite acetaldehyde affect N-glycan chain synthesis in the Golgi apparatus. Thus, Stibler and Borg[150] measured diminished activities of galactosyl-transferase and N-acetylglucosaminyl-transferase in serum from alcoholics and the addition of *in vitro* acetaldehyde to these samples further decreased the enzyme activities.

CDT glycoforms are diagnostic almost exclusively for alcohol abuse, in fact there are not diseases that cause the increase of CDT-glycoforms with the exception of Congenital Disorders of Glycosylation (CDG), a rare neurological diseases present in northern Europe. CDG, formerly named carbohydrate deficient glycoprotein syndrome, are rare hereditary disorders caused by mutations in the genes coding for enzymes involved in the biosynthesis of glycoproteins and other glycoconjugates. Stibler [145] investigated patients with CDG and all patients showed elevated CDT values. In spite of this, when CDT is assayed using an HPLC (High Performance Liquid Chromatography) or CE (Capillary Electrophoresis), asymptomatic CDG carriership can be detected.

The diagnosis of chronic alcohol abuse is difficult because many subjects deny or minimize the problem and diagnostic parameters with both high sensitivity and specificity are lacking. From both a medical and a social perspective there is a need for early detection and correct judgement of excessive alcohol drinking.[151] Indications in the field of forensic medicine are, for instance, drunken drivers, license reapplication,[152] especially of professional drivers, or *post-mortem* diagnosis of alcoholism from vitreous humor.[153] In occupational medicine, compelling indications are delegation to working sites where the impact of chronic alcohol consumption may have fatal consequences, such as aviation, working at high altitudes, handling of dangerous materials, and driving of potentially endangering machines. Among the currently used laboratory tests, CDT is regarded as the parameter with the highest diagnostic efficiency.[154] At least eight countries in Europe use biomarkers routinely as part of the clinical evaluation for drunk drivers and to assess if the person is abstaining.[155] In Switzerland, Italy, and Austria, repeat offenders are sent to therapy and biological markers, including CDT, are measured quarterly for an entire year to monitor alcohol abstinence. After 1 year, if treatment was successful and biomarkers were kept in check, the driver's license is reinstated. The mean half-life of CDT is approximately 17±4 days.[156] Due to the rapid

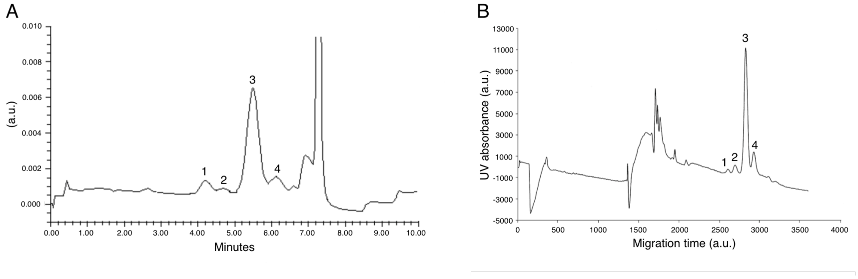


Figure 3.1.3: HPLC chromatogram (panel A) and CE electropherogram (panel B) of CDT positive samples. 1= disialo-Tf, 2=trisialo-Tf, 3=tetrasialo-Tf and 4=pentasialo-Tf.

decline of these modified Tfs values compared with indicative liver enzymes and mean corpuscular volume, they are a valuable parameter in the early stage of alcohol withdrawal. If liver enzymes are still elevated, a decreasing CDT value supports the presumption of abstinence.[157]

Morgan [158] investigated the potential impact of CDT values on the decision of license reapplication in “high-risk offenders”. CDT values were useful in 70% of the cases. In 8% these values would have produced a change in the reapplication decision and in another 22% might have produced confusion. Iffland [159] investigated the significance of CDT values in 534 drivers suspected of having driven impaired. Approximately 20–25% of these drivers had serious alcohol problems. In the entire group, 55% of the subjects showed values exceeding 20 U/L (1 U correspond to 1 mg Tf), and approximately 30% of the group exceeding 30 U/L. CDT was proposed as a valuable parameter in combination with the commonly used parameters for detecting drivers with chronic alcohol abuse.[155]

CDT quantitative determination has mostly been carried out with commercial immunoassay methods which include a sample extraction and purification based on disposable ion-exchange cartridges. At present, CDT analysis can be performed also with alternative methods, based on instrumental liquid separation techniques, such as HPLC and CE (see Figure 3.1.3).

These methods, in comparison to immunoassays, offer clear advantages in terms of accuracy and precision, because of a more sophisticated separation of Tf glycoforms, as clearly demonstrated by different studies.[146, 148, 155] With HPLC serum samples are Fe^{III} saturated before injection into the HPLC system. Separation is performed on an ion exchange chromatography column using salt gradient; quantification relies on the selective absorbance of the Tf-Fe^{III} complex at 460 nm. The relative amount of each glycoform is calculated as a percentage of the area under the curve, using baseline integration.[160] Detection of Tf using HPLC with absorbance at 460 nm is highly specific for the protein–Fe^{III} complex. Therefore, HPLC is a robust method against interference by other serum proteins. This technique remains however less sensitive than CE. The high efficiency of CE enables excellent separation of CDT from normal Tf and well separates also asialo-Tf, never present in non-alcohol dependent subject, and non detectable by HPLC.[154] The analytical procedure works as follows: serum, previously saturated by incubation for 30 min with $FeCl_3$ and $NaHCO_3$, is directly infused by hydrodynamic injection in the electropherograph. Separation is performed in an uncoated fused silica capillary by applying a voltage of 15 kV and the Tf glycoforms are then quantified. Tf-Fe^{III} complexes have their specific excitation wavelength (460 nm) but, because in the capillary is injected a small quantities of sample, it is used a wavelength of 200/215 nm (UV) where all proteins absorb.[161] Therefore, CE with UV detection is less specific for Tf than HPLC, even if much more sensitive. For this reason, developing an improved CE assay that would allow a more specific CDT detection has become one of the challenges in this scientific field.

A fluorescent probe is a molecule employed for the detection of biomolecules. Generally, fluorescent labeling consists in a reactive derivative of a fluorescent molecule known as a fluorophore. The fluorophore selectively binds (chemically or biologically) to a specific region or functional group on the target molecule, allowing the detection of the analyte by fluorescent spectroscopy. Some lanthanide (Eu^{III}, Tb^{III}, Sm^{III}, and Dy^{III}) complexes are known to be luminescent.[162] Compared with organic fluorescent compounds, the fluorescence of lanthanide chelates has several special different properties:

- the lifetime of the lanthanide chelates is very long; that of Eu^{III} and Tb^{III} chelates usually ranges from several hundred microseconds to more than one millisecond, and that of Sm^{III} and Dy^{III}, 10 to 100 microseconds;
- stokes shifts are very large;
- the emission profiles are sharp, having a full width at half maximum of only $\sim$10 nm.

In the last 20 years, lanthanide chelates have been successfully developed as fluorescence labels for highly sensitive detection of various biological molecules in time-resolved fluorometry.[162] Lanthanide fluorescence labels have been used in DNA hybridization assay, HPLC, CE and other bioassays, and have shown great improvement of the sensitivity compared to the conventional fluorometry using organic fluorescent labels.[163] The high sensitivity or detectability is due to easy distinction of the specific fluorescence signal of the lanthanide labels from background signals present in most biological samples. Time-resolved fluorometry also obviates the problems associated with light scattering of the optical components.[162]

Transferrin is able to bind Tb^{III} whose luminescence can be sensitised by energy transfer from the intrinsic protein fluorophores of the aromatic amino acids, tyrosine, tryptophan, and phenylalanine.[164] The efficiency of this transfer in Tb^{III}–protein complexes depends on the distance between the sensitising amino acid and the co-ordinated metal ion. However, in transferrin, where Tb^{III} is directly coordinated to a tyrosinate residue, highly efficient sensitisation of the lanthanide ion emission takes place.[164]

This research project has the purpose of characterize Tf-Tb^{III} complex and determine which experimental conditions make this bond more stable in order to achieve a deeper comprehension and use this knowledge to improve CDT detection. Different spectroscopies have been used to analyze different characteristics of the Tf-Tb^{III} complex to find the condition that allows a more stable Tf-Tb^{III} binding. To get more information on the binding mode, a method that allows the determination of the molecular model of the Tf interaction site with metal ions has been optimized through computational techniques. This method can be used, in future studies, for the determination of Tb^{III} environment surroundings in Tf active sites comparing experimental and computed spectra.

3.1.1.2 Methods

Experimental section

$TbCl_3$ was purchased from Sigma Chemical CO. A stock solution was prepared by dissolving weighed samples of the chloride in a small volume of 100 mM ammonium acetate, pH 8.1. The solution was diluted to volume with 100 mM ammonium acetate and the final pH was 8.1. Apo-Tf and holo-Tf were purchased from Sigma Chemical CO. Solutions of Tf were prepared by dissolving the protein into a ammonium acetate buffer, pH 8.1. The protein concentration was determined by spectrophotometric method.

UV-Vis absorption spectra were measured on a UV-vis spectrophotometer, using a cuvette with a 1-cm pathlength. The spectra was obtained scanning the wavelength ranging from 200 to 700 nm. The UV-Vis spectra of 14.5 µM apo-Tf and 10 µM holo-Tf dissolved in 100 mM ammonium acetate, 25 mM ammonium bicarbonate, pH 8.1 were performed. For apo-Tf, $TbCl_3$ was gradually added to the solution. At each addition the spectrum was recorded after 10 minutes in order to allow the formation of the protein-Tb^{III} complex. The concentrations of $TbCl_3$ tested were 10, 20, 50, 100, 200 µM.

Circular dichroism was used to study the effect of ammonium bicarbonate, the binding specificity of Tb^{III}, the Tb^{III} titration and the pH titration. CD spectra were recorded on a J-815 spectropolarimeter (JASCO Corporation, Tokyo, Japan) equipped with a Peltier system to control the sample temperature (set at 25 °C). The spectra were acquired in the near-UV region (scanning the wavelength range from 240 nm to 320 nm) employing a 1-cm pathlength cuvette, a data pitch of 0.1 nm, and a scan speed of 20 nm/min. All the spectra have been subtracted by the signal of the buffer alone. For pH titration 20 µM apo-Tf/holo-Tf were dissolved in a 100 mM ammonium acetate, 25 mM ammonium bicarbonate, pH 8.1 buffer and saturated with $TbCl_3$ (final concentration of 1500 µM). Ammonium hydroxide was added to the solution until the pH of the initial solution was 9.5. The sample has been left standing for ten minutes in order to allow the formation of the protein-Tb^{III} complex before the record of near-UV CD spectrum. Formic acid was gradually added to the solution; for each addition the spectrum was recorded after 3 minutes. The pH tested were between 3.5 and 9.5. To perform a titration with $TbCl_3$, this latter was gradually added to the solution. At each addition the spectrum was recorded after 10 minutes in order to allow the formation of the protein-Tb complex. The $TbCl_3$ concentrations tested were 10, 20, 50, 100, 200, 500, 1000, 1500 µM and the apo-Tf concentrations tested were 10, 20 µM.

A Vis-CD (400-700 nm) spectra was performed dissolving 200 µM of holo-Tf in MilliQ water in order to be compared with the simulated one (*vide infra*).

Fluorescence spectroscopy was used in order to study the effect of ammonium bicarbonate and the binding specificity of Tb^{III} and the Tb^{III} titration. Fluorescence spectra were recorded on a Cary Eclipse instrument (Varian, Palo Alto, CA, USA), exciting the sample at 295 nm (wavelength at which tryptophan absorbs) and monitoring emission in the range 310–575 nm. Quartz cuvettes of 1-cm pathlength were employed. Data were collected at 25°C. Raw data were corrected by subtracting the spectrum of the solvent. Fluorescence intensity is expressed in arbitrary union with the same scale in all experiments. For titration of apo-Tf *plus* Tb^{III} with ammonium bicarbonate 20 µM apo-Tf was dissolved in a 100 mM ammonium acetate, pH 8.1 buffer to which was added $TbCl_3$(final concentration of 100 µM). The solution had to be left standing for ten minutes in order to allow the formation of the protein-Tb^{III} complex. The emission spectrum of the sample was performed and the ammonium bicarbonate was gradually added to the solution: for each addition the spectrum was recorded after 5 minutes. pH was kept constant at 8.1 throughout the experiment. The concentrations of ammonium bicarbonate tested were 5, 10,15, 20, 25, 30, 35 mM. For Tfs titrations with a solution of $TbCl_3$ 10 µM holo-Tf was dissolved in a 100 mM ammonium acetate, 10 mM ammonium bicarbonate, pH 8.1 buffer. The concentrations of $TbCl_3$ tested were 100 and 200 µM. The same experiment was realized using apo-Tf dissolved in 25 mM ammonium bicarbonate. In this case the concentrations of $TbCl_3$ tested were 10, 20, 50, 100, 200, 500, 1000, 1500 µM. Apo-Tf concentrations tested were 5, 10, 20 µM.

CD, fluorescence and absorbance signals at the metal absorbing wavelength were plotted versus the $TbCl_3$ concentration and then fitted with a logistic equation (see eq 3.1.5, originlab, version8.0; OriginLab Corporation, Northampton, MA, USA) in order to obtain the titration curve.

$$\begin{cases} \frac{[L]_b}{[P]_t} = \frac{n[L]}{K_d+[L]} \\ [L]_t = [L] + [L]_b \end{cases} \tag{3.1.1}$$

$$\frac{[L]_b}{[P]_t} = \frac{n([L]_t - [L]_b)}{K_d + ([L]_t - [L]_b)} \tag{3.1.2}$$

$$K_d[L]_b + [L]_b[L]_t - [L]_b^2 = n[P]_t[L]_t - n[P]_t[L]_b \tag{3.1.3}$$

$$[L]_b^2 - (K_d + [L]_t + n[P]_t)[L]_b + n[P]_t[L]_t = 0 \tag{3.1.4}$$

$$[L]_b = \frac{K_d + [L]_t + n[P]_t - \sqrt{(K_d + [L]_t + n[P]_t)^2 - 4n[P]_t[L]_t}}{2} \tag{3.1.5}$$

This equation is valid for proteins having two binding-sites identical and indipendent, as previous studies suggest.

Computational section

Time-dependent Density Funtional Theory (TD-DFT) excited state energy was computed using TURBOMOLE suite of programs.[57] This computational setting [165] provides ground state geometry parameters in good agreement with experimental X-ray values, and a reasonable picture for excited state PES properties.

Atomic coordinates of the starting structure were obtained from the X-ray geometry of human Tf (PDB ID: 3V83).[166] Only Asp-63, Tyr-95, Tyr-188 and His-249 of N-lobe, in addiction to the carbonate ion and $\mathrm{Fe^{III}}$ center, were considered for DFT calculations. The residues were truncated at C_α and saturated with hydrogens. During geometry optimizations, C_α have been constrained to the crystallographic positions, in order to avoid unrealistic distortions at the boundary of the model. This starting structure was used to validate the level of theory. As reported previuosly,[167] $\mathrm{Fe^{III}}$ was considered in high-spin state ($S = 5/2$). The choice of TZVP basis set [59] and B3LYP hybrid functional,[101, 102] in combination with COSMO approach [85, 86] to model the protein environment polarizing effect (ε=4), has proved the most reliable.

CD spectra were computed using excitation energies, oscillator and rotatory strengths computed at TD-DFT level, using B3LYP in gas-phase. Computed spectra were obtained through a linear superimpositions of gaussian functions, centered at the TD-DFT excitation energies with the maximum values determined by the rotatory strengths. The full width at half maximum was set to 50 nm.

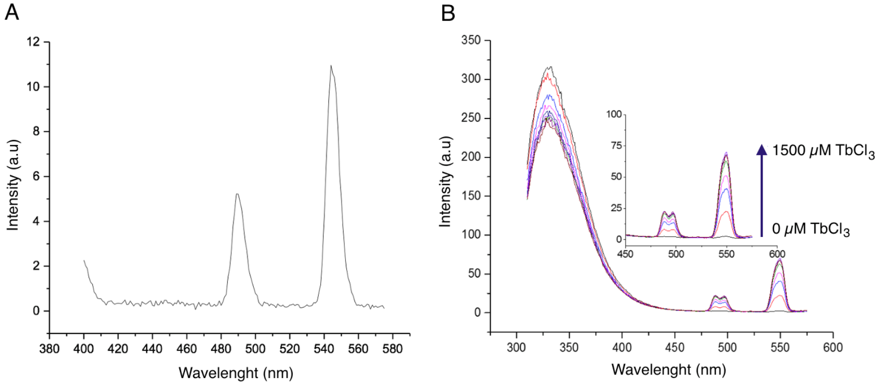


Figure 3.1.4: Apo-Tf titration with TbCl$_3$ by fluorescence. Panel A: fluorescence emission spectra of a 100 mM TbCl$_3$ solution (excitation is at 351 nm). Panel B: fluorescence emission spectra of apo-Tf titration with TbCl$_3$ (10, 20, 50, 100, 200, 500, 1000, 1500 µM; excitation is at 295 nm).

3.1.1.3 Results

Experimental section

Preliminary analysis were performed by fluorescence spectroscopy to determinate emission properties of Tb^{III} in absence and presence of Tf (Figure 3.1.4). A terbium chloride solution excited at 295 nm doesn't emit fluorescence, as is possible to see in Figure 3.1.4, panel A. The addition of TbCl$_3$ to a constant concentration of apo-Tf allowed to observe an intense fluorescence signal at 295 nm. Tf-Tb^{III} complex has a typical profile with a strong peak around 549 nm and another weak peak around 490 nm, wavelengths corresponding to the Tb^{III} electronic transition (Figure 3.1.4, panel B). Following the signal intensity at 490 nm it is possible to see that saturation was obtained at ca. 1000 µM. For this reason, in the following experiments, apo-Tf was incubated for 10 minutes with a TbCl$_3$ concentration of 1500 µM to obtain Tb^{III} saturated-Tf. Different observations, among which the non-fluorescence of Tb^{III} at 295 nm, strongly suggest that the fluorescence signal at 549 nm is due to Tf-Tb^{III} complex formation. First of all, a terbium chloride solution excited at 351 nm, wavelength at which Trp emits, shows no signal in the range of concentrations used in the titration, while at higher concentrations (100 mM TbCl$_3$) shows a fluorescence signal. In addition, the peak profile at 490 nm is sensitive to the surrondings chirality and its splitting can be a proof of the interaction with a chiral environment (as the protein environment).[168] These evidences also suggest that FRET (Fluorescence Resonance Electron Transfer) [169] occured in our experiments due to a partial overlap between the Trp emission and the terbium excitation spectra, and the location of tryptophans at a distance lower than Foster distance from $\mathrm{Tb^{III}}$ (assuming that $\mathrm{Tb^{III}}$ is bound in the same pocket of $\mathrm{Fe^{III}}$).

The stoichiometry of Tf-$\mathrm{Tb^{III}}$ complex was determined in collaboration with the laboratory of Prof. Rita Grandori at University of Milano-Bicocca. A nano-electrospray-ionization mass spectrometry (nano-ESI–MS) analysis was performed, and revealed that the binding stoichiometry of $\mathrm{Tb^{III}}$ to Tf is 2:1, as in $\mathrm{Fe^{III}}$ complexes. In addition, no signal corresponding to higher stoichiometry are detected even at very large molar excesses of the metal. These evidences taken together are a strong indication of the binding specificity of $\mathrm{Tb^{III}}$ for Tf.

To evaluate the conformational properties of both apo- and metal-saturated forms of the protein have been employed (Figure 3.1.5). Near-UV CD spectroscopy is often used to probe tertiary structure of proteins and their complexes, as the intensities of the aromatic residue bands (phenylalanine at 255-270 nm, tyrosine at 280 nm and tryptophan at 290 nm) depend on a unique orientation of these chromophores, as determined by the higher order structure of the protein. The near-UV CD spectra of $\mathrm{Tb^{III}}$- and $\mathrm{Fe^{III}}$-saturated Tf exhibit very significant deviations from the spectrum of the apo-protein: a strong positive band above 300 nm in the $\mathrm{Fe^{III}}$-Tf spectrum and a negative band below 260 nm in the Tf-$\mathrm{Tb^{III}}$ spectrum (Figure 3.1.5, panel A). Thanks to these variations is possible to detect the formation of complexes. Most importantly, the presence of $\mathrm{Tb^{III}}$ and $\mathrm{Fe^{III}}$ does not strongly alter the shapes and intensities of the bands due to contribution of aromatic (260-290 nm), indicating very little changes in the tertiary structure of Tf upon metal binding. At the same time, neither $\mathrm{Tb^{III}}$ nor $\mathrm{Fe^{III}}$ induce any noticeable change in the far-UV CD spectra of Tf, consistent with the notion that the protein secondary structure is not very altered upon metal binding (Figure 3.1.5, panel B). Therefore, through far-UV CD, unlike near-UV CD, the complex detection is not possible. The UV-Vis absorpion spectra shown in Figure 3.1.5 panel B, permit to detect both the Tf-metal complexes. Tf-$\mathrm{Tb^{III}}$ spectrum has one strong peak around 245 nm, which arises from perturbation of the aromatic ring of the active site tyrosines, and another weak peak around 295 nm while Tf-$\mathrm{Fe^{III}}$ spectrum has

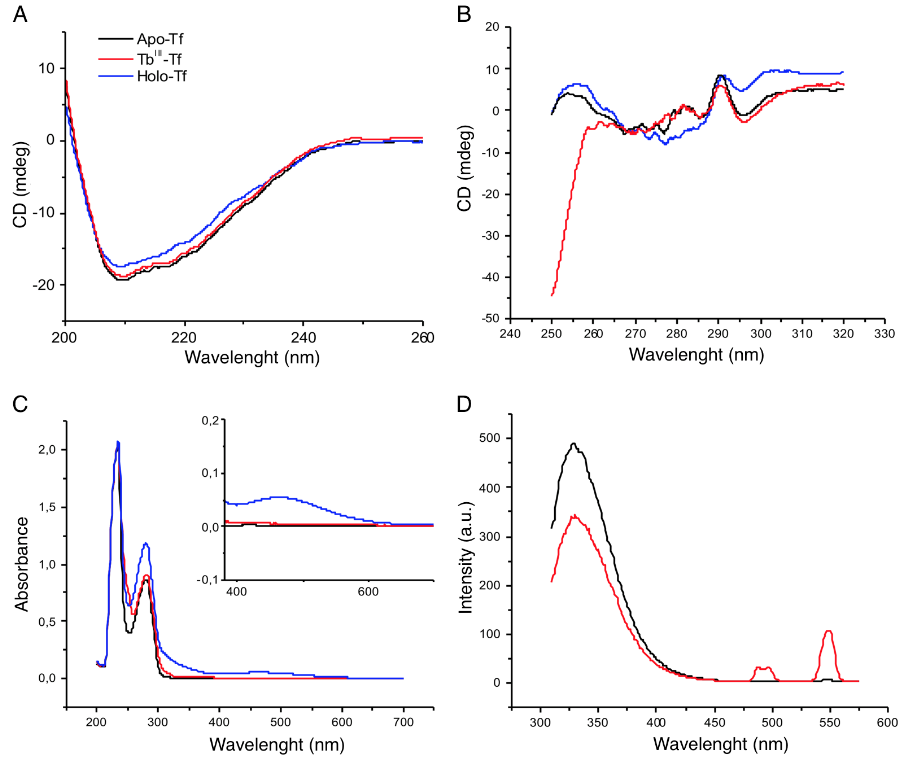


Figure 3.1.5: Spectroscopic characterization of protein metal complexes. Spectra of 20 µM apo-Tf (black), holo-Tf (blue) and Tb^{III}-Tf obtained by different techniques. Panel A: near-UV CD; panel B: far-UV CD; panel C: UV-vis absorbance; panel D: fluorescence spectroscopy.

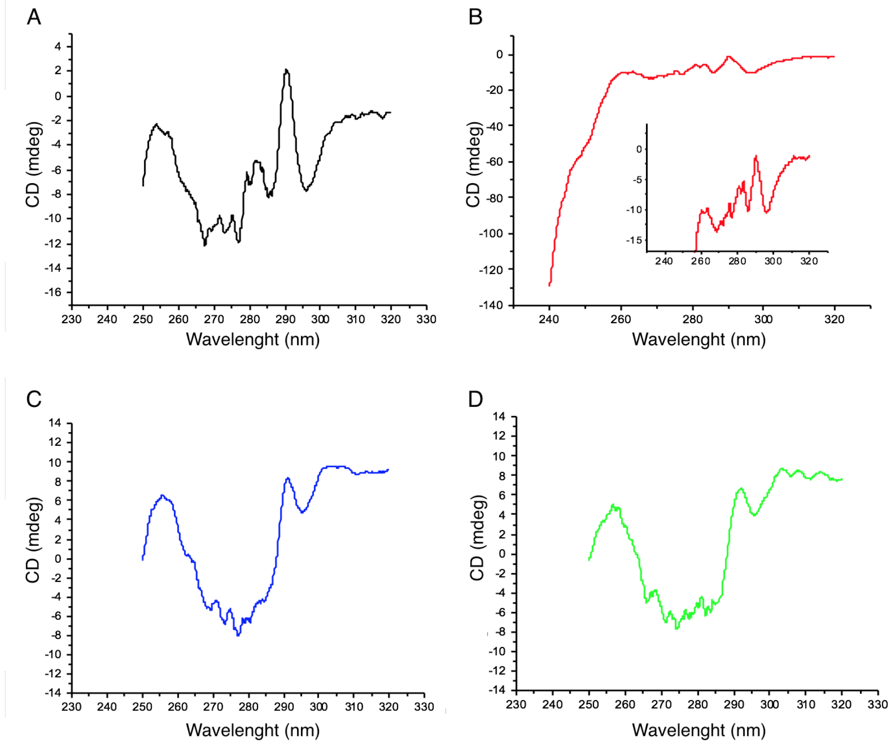


Figure 3.1.6: $\mathrm{Fe^{III}}$ and $\mathrm{Tb^{III}}$ competition by near-UV CD. Near-UV CD spectra of 20 µM apo-Tf (panel A), $\mathrm{Tb^{III}}$ saturated-apo-Tf (1500 µM $\mathrm{TbCl_3}$, panel B), holo-Tf (panel C) and holo-Tf in the presence of 1500 µM $\mathrm{TbCl_3}$(panel D).

a peak at 465 nm. By fluorescence instead (Figure 3.1.5, panel D), following the signal intensity at 490 nm it is possible to track the Tf-$\mathrm{Tb^{III}}$ complex formation, while as expected holo-Tf complex is not detectable.

To have a deeper comprehension of Tf-$\mathrm{Tb^{III}}$ binding process, holo-Tf titration with increasing concentrations of $\mathrm{TbCl_3}$ have been recorded in the presence of 25 mM ammonium bicarbonate, pH 8.1 by both CD and fluorescence. Unlike the $\mathrm{Tb^{III}}$ saturated-apo-Tf CD spectrum, which shows variations from the apo-Tf spectrum under 260 nm, holo-Tf in the presence of $\mathrm{TbCl_3}$ (Figure 3.1.6, panel C) spectrum is perfectly superimposable to holo-Tf CD spectrum (Figure 3.1.6, panel D). The fluorescence spectrum of holo-Tf, even in the presence of a 10-fold $\mathrm{Tb^{III}}$ molar excess, relative to holo-Tf, does not present the typical Tf-$\mathrm{Tb^{III}}$ profile around 475/575 nm (Figure 3.1.7, panel D). The fluorescence

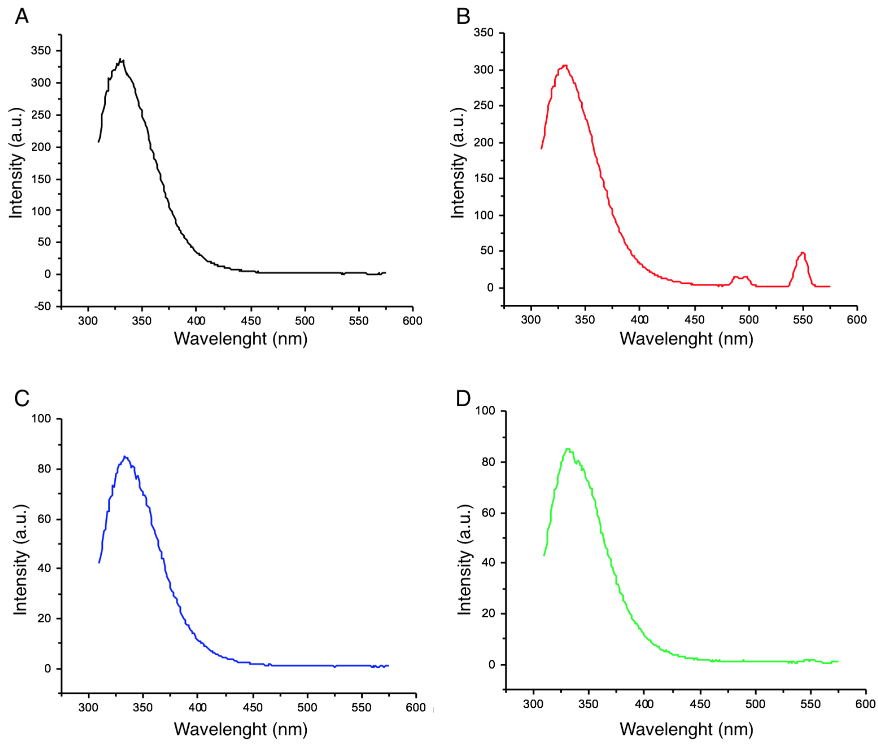


Figure 3.1.7: Fe^{III} and Tb^{III} competition by fluorescence. Emission fluorescence spectra of 20 µM apo-Tf (panel A), Tb^{III} saturated-apo-Tf (1500 µM TbCl$_3$, panel B), holo-Tf (panel C) and holo-Tf in the presence of 200 mM TbCl$_3$(panel D). Excitation is at 295 nm.

profile is highly comparable to the normal profile of holo-Tf, not incubated with $TbCl_3$. All together these results indicate direct competition between Fe^{III} and Tb^{III} for the same binding sites and remarkable higher affinity of Fe^{III} relative to Tb^{III}.

In order to quantify the strength of Tf-Tb^{III} interaction, apo-Tf titrations by increasing $TbCl_3$ concentrations in the presence of 25 mM ammonium bicarbonate were carried out using different spectroscopic methods: CD, fluorescence and UV-Vis absorbtion. The intensity of Tb^{III} specific bands were plotted versus $TbCl_3$ concentration and then fitted with a logistic equation in order to obtain the titration curves (Figure 3.1.8).

In absorption spectra (Figure 3.1.8, panel A), peaks at 245 and 295 nm arise while in fluorescence spectra (Figure 3.1.8, panel B) signals at 495 and 549 nm increase in intensity. In near-UV CD spectra the band below 270 nm reduces its intensity (Figure 3.1.8, panel C). In UV-Vis spectra the $TbCl_3$ concentrations shown is limited to 200 µM because higher concentrations gave poor quality spectra. The K_d values obtained for Tb^{III} by these techniques approximately range within the same order of magnitude (ca. 10^{-5} M, Figure 3.1.8, panel B, D, F) and are in agreement with the ones already reported by Abdollahi and colleagues.[170] Moreover the K_d obtained have proved to be independent of the protein concentration used (data not shown). These values are much higher than the reported K_d for Fe^{III} (10^{-22} M),[144] consistent with the results discussed above and with Fe^{III} being the natural ligand of Tf. As reported by the study of Xiao,[171] transport proteins responsible for chaperoning essential metals to their cellular destinations appear to bind the metal ions with high selectivity and affinity ($K_d < 10^{-7}$ M) to guarantee the biological function and to minimize the toxic effects of other free metal ions.

The model used to obtain the titration curves is based on the assumption that the two binding sites are equal and independent. For this reason, it would be appropriate to carry out future experiments using different models (see equation 3.1.5 in the methods section).

As said in the introduction section, both Tf binding sites, require a bidentate ligand carrying two negatively charged oxygen atoms to further stabilize Fe^{III} complex. This role is normally played by one molecule of carbonate ion (CO_3^{2-}). For this reason, all the experiments above were performed in the presence of ammonium bicarbonate. Thus the effect of the presence of the

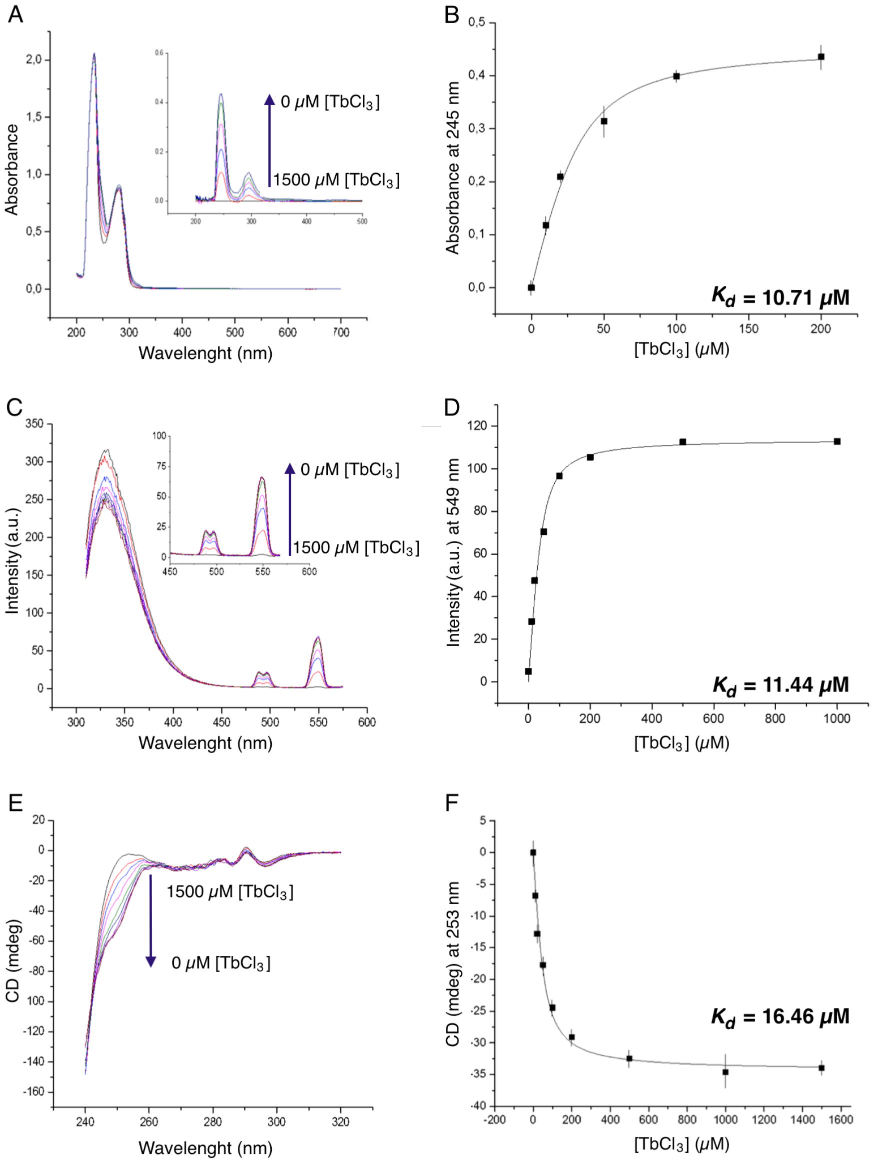


Figure 3.1.8: Raw UV-Vis (panel A), fluorescence (panel C) and CD (panel E) spectra and respective data fitting (panels B, D, F) obtained through the titrations of 20 μM apo-Tf by increasing $TbCl_3$ concentrations. The fitting of the curve and K_d values were obtained through a logistic equation based on two identical and indipendent binding sites.

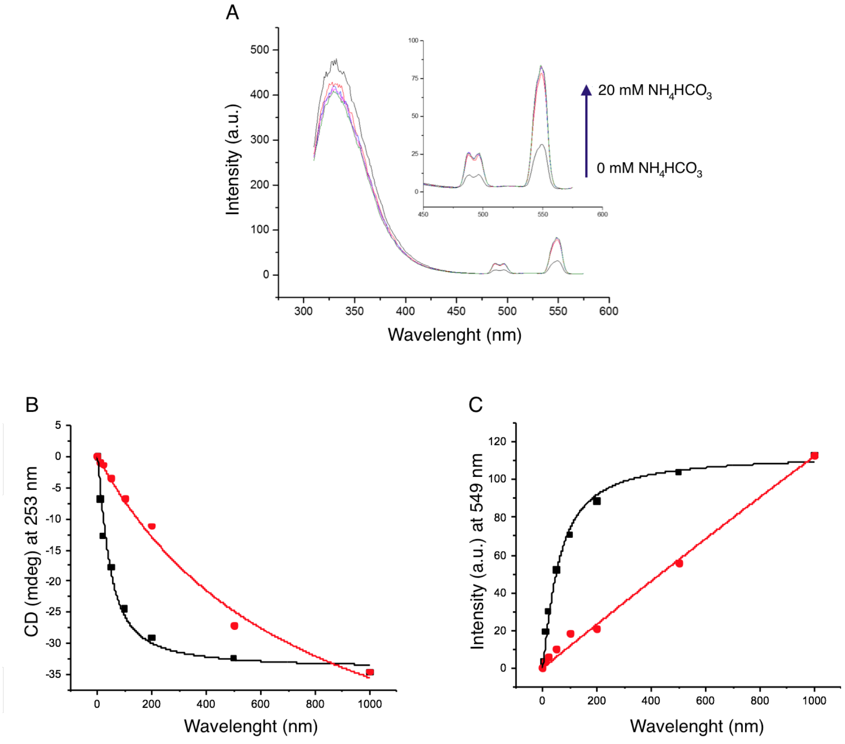


Figure 3.1.9: Effect of ammonium bicarbonate on Tf-Tb^{III} complex formation. Panel A: 20 µM Tf-Tb^{III} titration with increasing ammonium bicarbonate concentrations obteined by fluorescence spectroscopy. Excitation at 295 nm. Titration curves (near-UV CD, panel B and fluorescence, panel C) obtained through the addition of increasing concentration of TbCl$_3$ to 20 µM apo-Tf in the presence (black) and absence (red) of 10mM NH$_4$HCO$_3$.

synergistic anion was also tested on Tf-Tb^{III} complexes. A fluorescence experiment was performed with fixed protein (20 µM) and terbium (100 mM) concentrations and increasing ammonium bicarbonate (NH$_4$HCO$_3$ concentrations (Figure 3.1.9, panel A). A progressive increase of the fluorescence signal intensity can be observed. The process displays a saturation behavior, without further increase in signal intensity, above a ammonium bicarbonate concentration of 10 mM. This result suggests that ammonium bicarbonate assists the binding of Tb^{III} to the protein. The presence of the anion leads to a greater amount of complexes formed for each given protein-ligand mixture. In order to

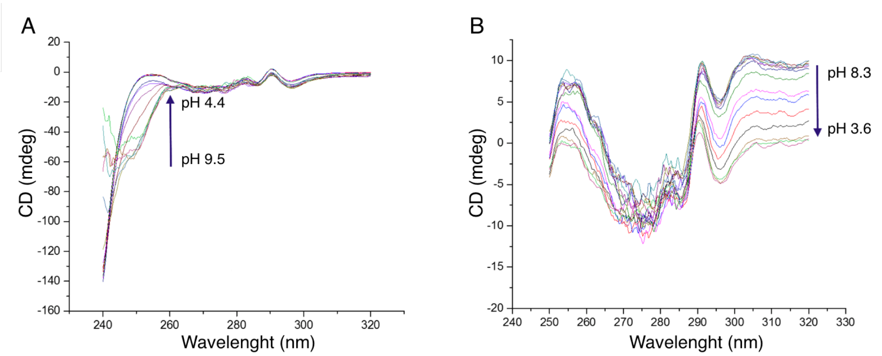


Figure 3.1.10: pH stability comparison by near-UV CD. Titration of 20 µM Tb^{III} saturated-Tf (panel A) and holo-Tf (panel B) has been performed by serial additions of 10% formic acid.

extimated the effect of bicarbonate on Tf-Tb^{III} affinity, titrations of apo-Tf by increasing concentrations of terbium chloride were carried out in the presence (10 mM) and in absence of ammonium bicarbonate and monitored by near-UV CD or fluorescence. Titration curves are shown in Figure 3.1.9 (near-UV CD, panel B and fluorescence, panel C).

The profiles indicate much weaker binding in the absence of ammonium bicarbonate and no formation of a plateau in the tested ligand concentration range, different from the profiles in the presence of the synergic anion. The K_d of Tf for Tb^{III} in the absence of ammonium bicarbonate increases of approximately two orders of magnitude (ca. 1200 µM).

The effect of pH in the range of 3.5-9.5 on the stability of Tb^{III} saturated-Tf an holo-Tf have been tested (Figure 3.1.10, panel A and panel B respectively). Near-UV spectra of holo-Tf and Tb^{III}saturated-apo-Tf have been performed in the presence of ammonium bicarbonate at variable pH. The spectra show a variation in correspondence of the bands previously identified, results of the formation of the metal-protein complex. As can be seen from the comparison between Figure 3.1.10, panel B and Figure 3.1.6, panel A, the holo-Tf and the apo-Tf CD spectra became almost coincident around pH 4; the protein has completely lost Fe^{III} at that pH. Based on signal variation at 296 nm for Fe^{III} and 253 nm for Tb^{III}, the transition profile for the two complexes have been derived (Figure 3.1.11). As it is possible to see in Figure 3.1.11, a shift between the two titration curves occurred. Specifically, the holo transition takes place in

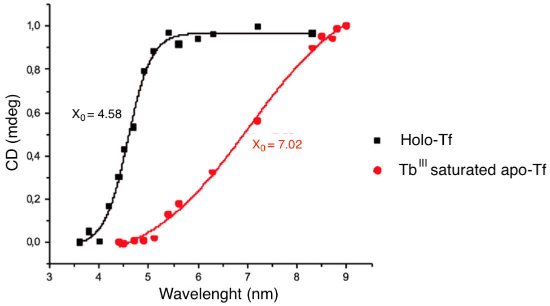


Figure 3.1.11: Titration curves (CD intensities of the metal specific bands plotted versus the pH values) of holo-Tf (black) and Tb^{III}-saturated apo-Tf (red). X_0 is the inflection point of the curve which represents the pH at which half protein has lost metal.

a narrower range (pH range 4-5.5) and towards lower values than Tb^{III} complex transition (pH range 5-9).

This results indicate that Tf-Tb^{III} complexes are more sensitive to pH variation and may suggest that the coordinating sphere of the metal centre could be slightly different for the two metals. It seems likely that, in the case of Tb^{III}, amino acids with high pKa contribute to the complex stabilization. It could not be ruled out that the net charge of the protein also plays a role in stabilizing Tf-Tb^{III} complexes.

Computational section

Recent advances on calculation of electronic CD have greatly enhanced its value in determining absolut configuration of chiral molecules and metal centres. Quantum chemistry methods including TD-DFT can be used to calculate or predict the theorical CD spectra.[172]

To assess the feasibility and reliability of the TD-DFT method, the first calculation of CD spectrum has been performed on the native metal binding site of human Tf. The coordination geometry of Fe^{III}has been derived from the crystallographic structure of human Tf (PDB ID: 3V83).[166] The residues Asp-63, Tyr-95, Tyr-188 and His-249 of N-lobe, in addiction to the carbonate ion and Fe^{III} center, were considered for DFT calculations. The optimized geometry at B3LYP/TZVP level of theory considering a dielectric mimicking the protein environment is shown in Figure 3.1.12.

As shown in Figure 3.1.12 the structural variation between the simulated and the crystallographic structure is negligible. The calculated electron transitions and the simulated CD spectrum of human Tf binding site in the range 400-700 nm are shown in Figure 3.1.13 and Table 3.1 respectively.

Excitation	Transition	λ	f	R_{lenght}	%
1	137β→138β	643	$1.5{\cdot}10^{-3}$	12.7	67.7
2	137β→139β	621	$1.0{\cdot}10^{-3}$	12.9	71.3
3	137β→140β	606	$2.3{\cdot}10^{-3}$	-50.5	66.0
4	136β→138β	495	$1.7{\cdot}10^{-3}$	-49.2	73.5

Table 3.1: Key transitions and their related oscillator and rotatory strengths at B3LYP/TZVP level. λ: wavelenght in nm; f: oscillator strenght; R_{lenght}: rotatory strength in length form ($10^{-40}{\cdot}\text{erg}{\cdot}\text{cm}^3$); %: main 1-electron contribution percentage.

Visible CD spectroscopy is a very powerful technique to study metal-protein interaction and can risolve individual d-d electronic transitions as separated bands. CD spectra in the visible light region are only produced when a metal ion is in a chiral environment, thus, free metal ions in solution are not detected. This has the advantage of only observing the protein-bound metal and its optical activity attributed to configurational, conformational and vicinal effects. In this study both far- and near-UV spectra range ($<$ 400 nm) are not taken into consideration because here the signals are due to aromatic content in the protein environment. Since the geometry of the binding site of Tf has been modeled with two aromatic tyrosines (in addition to the other residues listed previously), in order to make a reasonable comparison the wavelength range considered is 400-700 nm.

The calculated spectrum is consistent with its experimental data in this region, although the maximum absorption is shifted toward the high wavelength region, i.e., around 495 nm in the calculated *vs* around 460 nm in the experimental spectrum. The molecular orbitals involved in key transitions for the calculated

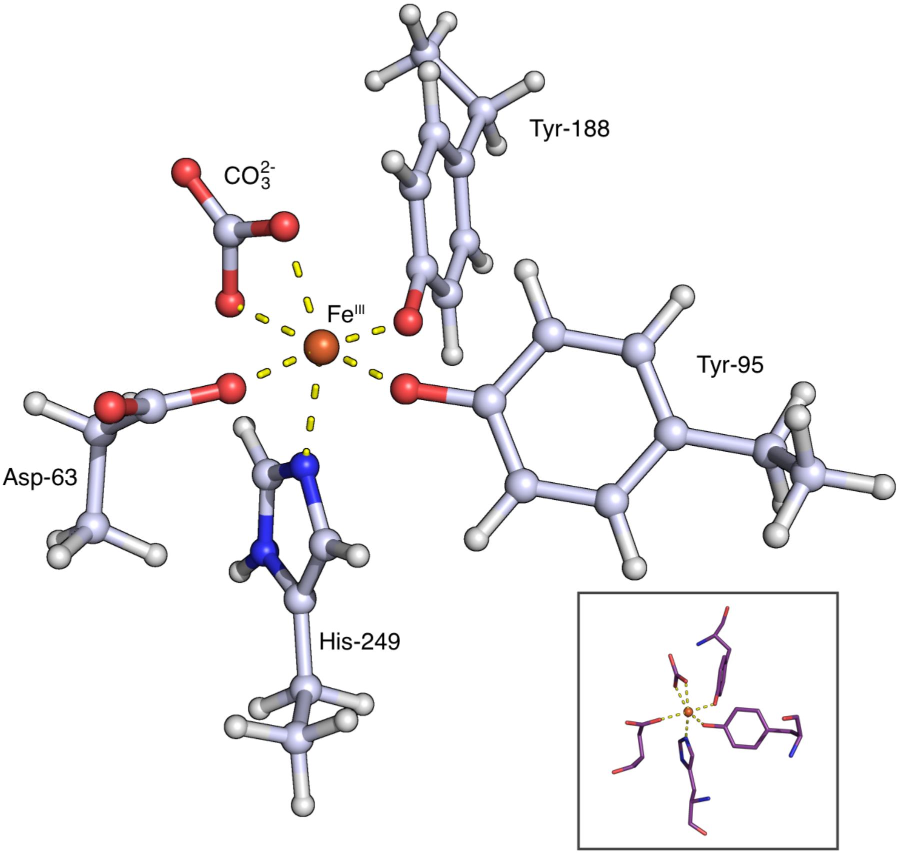


Figure 3.1.12: Optimized geometry at B3LYP/TZVP (ε=4) level. In the square is represented the crystallographic geometry of the metal binding site (PDB ID: 3V83, N-lobe).

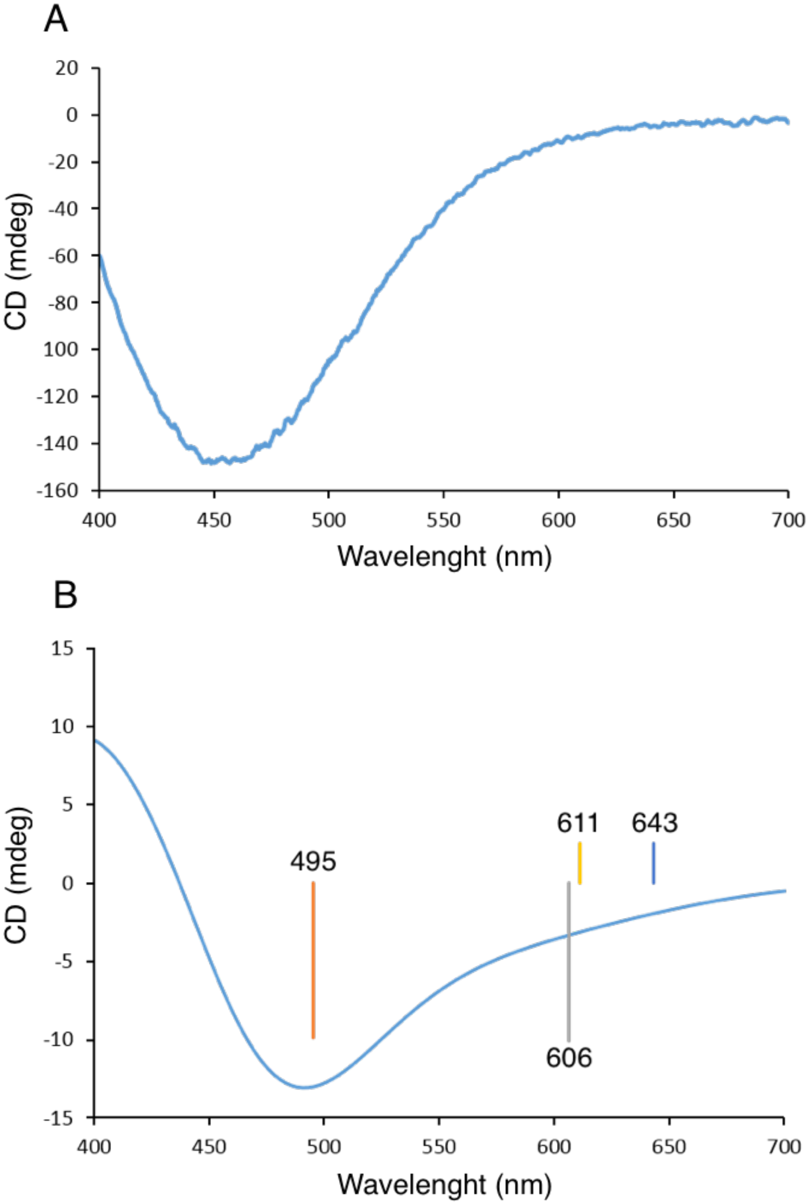


Figure 3.1.13: Comparison between experimental (panel A) and simulated (panel B) CD spectra in the range of 400-700 nm. The coloured bars represent the key electronic transitions and their intensity.

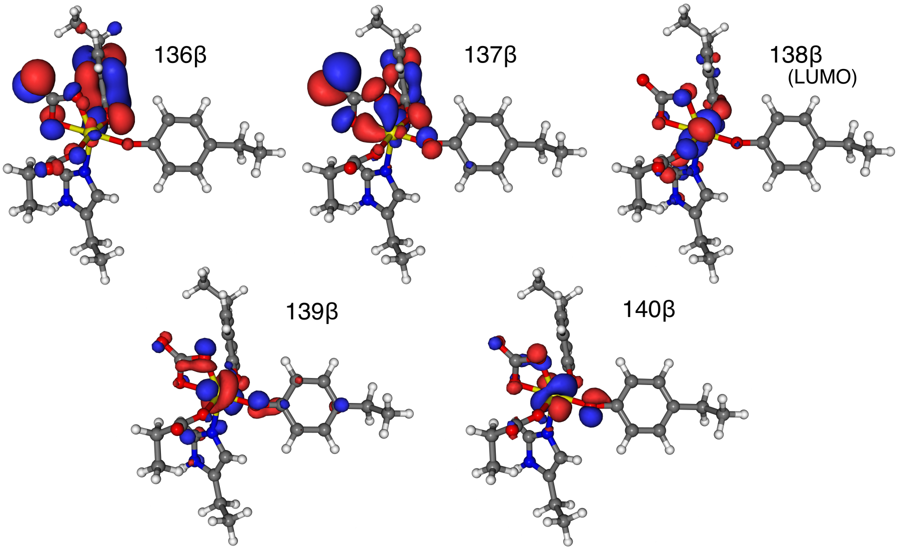


Figure 3.1.14: Molecular orbitals involved in key transitions in calculated CD spectra of holo-Tf at B3LYP/ TZVP level. Isosurface cut off value is 0.03.

spectra of the Tf metal binding site are enumerated in Table 3.1 and shown in Figure 3.1.14.

The major negative rotatory strenght at 606 nm results from the electronic transition from orbital 137β scattered on the aromatic ring of Tyr-188 and the sinergistic carbonate anion to the unoccupied orbital 140β located on the iron atom. The negative CD signal generated from this transition is compensated by the close positive signals at 621 nm and 643 nm, causing a flattening of this negative peak. The two positive signals originate from the transitions of the electron in orbital 137β to orbital 139β and 138β LUMO (Lower Unoccupied Molecular Orbital) respectively, both located on the central iron atom. The latter signal considered at 495 nm shows an highly negative rotatory strenght. The electronic transition takes place between the orbital 136β located mainly on the aromatic ring of Tyr-188 and to a small extent on the carbonate anion and LUMO orbital on the iron atom. Due to the reliability of the method in simulating CD spectra, this procedure can be used in future studies, for the determination of $\mathrm{Tb}^{\mathrm{III}}$ surroundings in Tf binding site.

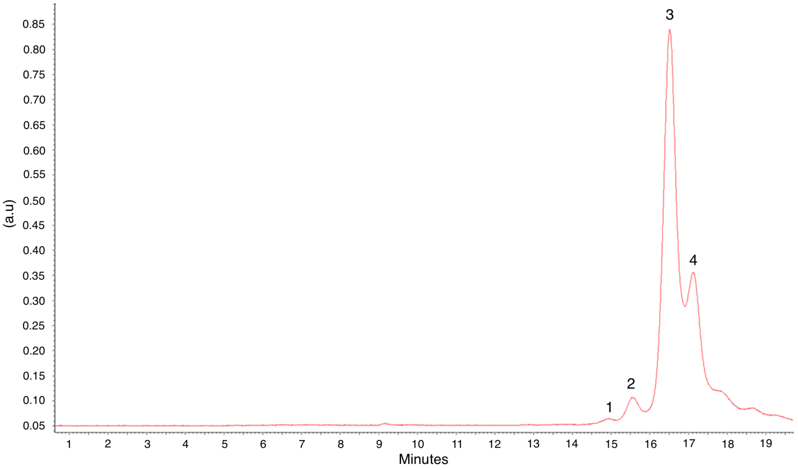


Figure 3.1.15: CE electropherogram of a positive serum sample after Tf-Tb^{III} fluorescent adduct formation.

3.1.1.4 Conclusions

CDT is currently the most specific and sensitive parameter for the diagnosis of chronic alcohol abuse. Analytical methods for CDT detection would benefit from the employment of a specific detector system. Thus, this work demonstrates the properties of the Tf-Tb^{III} complex that could be used as a fluorescent label during the detection.

The complex formation has been analyzed through different spectroscopic techniques. The results obtained reveal that Tb^{III} binds to Tf specifically and with a stoichiometry of 2:1, and its affinity for Tf is lower than Fe^{III} one. These data are also confirmed by K_d values: 10^{-22} M for Fe^{III} and 10^{-5} M for Tb^{III}. CO_3^{2-} anion is fundamental to assist Tf-Tb^{III} complex formation as it occurs for Fe^{III}. Future studies may be aimed at the analysis of other anions. In the case of Tf-Fe^{III}complexes, oxalate is particularly effective as a substitute synergistic anion.[173] For this reason, try it would be interesting to study the effects of the presence of oxalate anion also in Tf-Tb^{III} complexes. A method that allows the determination of the molecular model of the Tf interaction site with metal ions has been optimized by computational techniques. The procedure is based on the comparison of the holo-Tf experimental circular dichroism spectrum with its simulation obtained from a binding site crystal structure. To gain a wide comprehension of Tb molecular mechanism of binding, studies based on this computational method will be made for the determination of terbium surroundings in Tf active sites comparing experimental and computed spectra. All these data support the idea of using Tf-bound Tb^{III} as a fluorescent probe in CDT detection. Recently, in collaboration with the research group of Franco Tagliaro (University of Verona), the use of Tb^{III}-bound Tf in CE analysis has been tested (Figure 3.1.15).

From a comparison with Figure 3.1.3 (panel B) it is possible to see that the signal quality is significantly improved and now the detection system is specific and able to differentiate Tf from other proteins. Further studies will be anycase fundamental to optimize the protocol for a scale-up usage.

Chapter 4

Computational methods to support drug design: small ligands development and binding.

> *Mysteries do not lose their poetry when solved.*
> *Quite the contrary; the solution often turns out more beautiful*
> *than the puzzle and, in any case, when you have solved one mystery*
> *you uncover others, perhaps to inspire greater poetry.*
>
> *Richard Dawkins*

Medicinal chemistry research has successfully incorporated many molecular modeling methods, to study complex biological and chemical systems. The integration of computational and experimental strategies has been of great value in the identification and development of novel promising compound. Following the development of the first algorithms in the 1980s, molecular docking became an essential tool in drug discovery.[174] Abundantly used in modern drug design, it explores the conformations adopted by a small molecule within the binding site of biological macromolecule targets. This approach can also estimate the ligand-target binding free energy taking into account all chemical interactions involved in the recognition process.

The original concept of docking comes from the concept of *lock and key*, but the algorithms used to fit the *key* (the ligand) into the *lock* (the target) vary among the different programs.[175] In this step, called *conformational search*, structural parameters of the ligands such as torsional, translational and rotational degrees of freedom are incrementally modified. This task is performed by search alogirthms by applying systematic or stocastic research methods.[176, 177] *Systematic search* is promoted by slight variations of the structural parameters of the ligand, gradually changing its conformation.[178] The energy landscape of the conformational space is probed by the algorithm and, after a certain number of cycles, converges to the minimum energy solution corresponding to the most likely binding mode. Starting from different point of the energy landscape it is possible to overcome the drawback of converging in a local instead of a global minimum. *Stochastic methods* carry out the conformational search by randomly modifying the structural parameters of the ligands.[179] The algorithm generates ensembles of conformations and populates a wide range of energy landscapes. This strategy avoids trapping the final local solutions in a local minimum, increasing the probability of finding a global minimum. Since with this method there is an increase of the coverage of the energy landscape, the computational cost associated with this procedure is an important limitation.[179]

As already presented in Chapter 1, the *scoring function* allows to estimate the energy of a ligand conformation within the biological target. The energy variation, due to the formation of the ligand-protein complex, is given by the binding constant (K_d) and the gibbs free energy (ΔG).[180] Prediction of the binding energy is performed evaluating the physical-chemical phenomena involved in ligand-receptor binding, including intermolecular interaction, desolvation and entropic effects. Therefore the greater is the number of the parameters evaluated, the greater is the accuracy of the scoring function. On the other hand, there is an increase of the computational cost proportional to the number of variables included in the function, a flaw that reduces the performance of the docking algorithm. The ideal scoring function should offer a balance between accuracy and speed, a critical aspect when working with large ligand sets.

Here we present different docking studies conducted in collaboration with experimental groups of the Dept. of Biotechnology and Biosciences of the University of Milan-Bicocca. The chapter is divided in two section (Section 4.1 and Section 4.2), corresponding to the use of two different software DOCK 6 [181, 182] and Glide [183], respectively.

4.1 DOCK 6

DOCK 6 algorithm addresses rigid and flexible body docking using a geometric matching algorithm to superimpose the ligand into a negative image of the binding pocket and it works assisted by a series of accessory programs.

The starting point of all docking calculation is the identification of the potential binding site even if often, this site, is known *a priori*. To identify the active site, DOCK 6 uses a sets of overlapping spheres generated from the molecular surface of the receptor, in order to create a negative image of the surface invaginations of the target. The sphere centers are putative ligand atom positions. Their use is an attempt to limit the enormous number of possible orientations within the active site because the number of orientations of a ligand in free space is vast. The number of orientations possible from all sets of sphere-atom pairings is smaller but still large and cannot be generated and evaluated in a reasonable length of time. Each sphere is generated tangent to receptor surface points *i,j* with the center on the surface normal of point *i*. Spheres are calculated over the entire surface, producing approximatively one sphere per surface point. This representation is then filtered in order to keep only the largest sphere associated with each surface atom. The filtered set of spheres is then clustered using a single linkage algorithm in order to define different clusters representing the different invaginations in the target that are then ranked according to size.

A set of sphere is required a required input to generate the grid. The largest cluster is typically the ligand binding site of the receptor.

The grid generation is necessary for rapid score evaluation in DOCK. The energy scoring component of DOCK is based on the implementation of force field scoring. Force field scores are approximate molecular mechanics interaction energies, consisting of van der Waals and electrostatic components:

$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left(\frac{A_{ij}}{r_{ij}^{a}} - \frac{B_{ij}}{r_{ij}^{b}} + 332 \frac{q_i q_j}{D r_{ij}} \right) \tag{4.1.1}$$

where each term is a double sum over ligand atoms i and receptor atom j, where r_{ij} stands for the distance between protein atom i and ligand atom j, A_{ij} and B_{ij} are the van der Waals parameters, $a(=6)$ and b $(=12)$ respectively the attractive and the repulsive Lennard-Jones terms for van der Waals potentials, q_i and q_j are the atomic charges and D is the distance dependent dielectric constant.

Using the coordinates of an imaginary box built around the selected cluster of spheres the program pre-computes the energy potential for the active site at a specified grid spacing.

Flexible docking allows the ligand to be flexible and structurally rearrange in response to the receptor. According to the *anchor and grow* algorithm the largest rigid substructure of the ligand (the anchor) is identified and, in case it is an aromatic ring, all bonds are treated as rigid. If the molecule does not have a ring, the largest rigid segment is specified as the anchor. When an anchor is selected, then the molecule is re-divided into non-overlapping segments, which are arranged concentrically about the anchor segment.

The cluster of spheres selected as binding site is then used to locate the ligand; only the coordinates of the sphere centers are used to orient ligands. Sphere-sphere distances are compared to atom-atom distances. Sets of sphere atom pairs are generated in the following manner: sphere i is paired with atom I if and only if for every sphere j in the set

$$|d_{ij} - d_{IJ}| < \varepsilon \tag{4.1.2}$$

where d_{ij} is the distance between sphere i and sphere j, d_{IJ} is the distance between sphere I and sphere J and ε is a cut-off distance. In this stage of the Anchor and grow algorithm the anchor is rigidly oriented in the active site and its orientations are evaluated and optimized with a simplex minimizer. Score optimization allows the orientation and conformation to be adjusted to improve the score. The optimizer uses the downhill simplex algorithm,[184] which does not require evaluation of derivatives. The orientations are then ranked according to their score, spatially clustered by heavy atom RMSD and prioritized. Specifically, after score ranking, the top-ranked configuration is set aside and used as a reference configuration for the first round of pruning. RMSD between each candidate and the reference configuration is then computed. Each candidate is then evaluated for removal based on its rank and RMSD using the following inequality:

$$\frac{rank_i}{RMSD(i,0)} > Nc \tag{4.1.3}$$

where $rank_i$ is the ranking position of conformation i, $RMSD(i,0)$ is the calculated RMSD between conformation i and top-ranked conformation 0 and N_c is the clustering cut-off value.

Finally, in the growth stage the flexible layers of the ligand are build onto the best anchor orientations within the binding site of the receptor that will help restrict the sampling of ligand conformations.

Usually, a second DOCK 6 calculation with a different scoring function is used in the final minimization, in order to score and rank of the ligand conformations. AMBER score implements molecular mechanics implicit solvent simulations with the all-atom AMBER force-field for protein atoms and the general AMBER force field (GAFF) for ligand atoms. The interactions between the ligand and the receptor are represented by electrostatic and van der Waals energy terms, and the solvation energy is calculated using a Generalized Born solvation model. The AMBER score is calculated as

$$E_{binding} = E_{complex} - (E_{receptor} + E_{ligand}) \tag{4.1.4}$$

where $E_{complex}$, $E_{receptor}$ and E_{ligand} are respectively the internal energies of the complex, receptor and ligand as approssimated by the AMBER force field with the Generalized Born surface area term. The calculation of these three terms uses the same protocol: minimization with a conjugate gradient method, followed by a MD simulation (Langevin molecular dynamics at constant temperature), another minimization and a final energy evaluation. [181, 182]

With this methodology we performed the docking study of cannabigerol and a series of derivatives of phosphatidylinositol 3-phosphate, molecule able to exert pharmacological effects *in vitro* and *in vivo* on nociception and oncogenic transformation respectively.

4.1.1 CBG binding to α_2-adrenoceptors

4.1.1.1 Introduction

Cannabigerol (CBG) is a poorly characterized phytocannabinoid, which was first detected in *Cannabis sativa*, and found not to induce Δ9-tetra-hydrocannabinol-like psychopharmacological effects in vivo (4.1.1, panel C). A recent investigation [185] has provided the first evidence that CBG is able to bind G-protein coupled receptors, and in particular activate α_2-adrenergic receptors

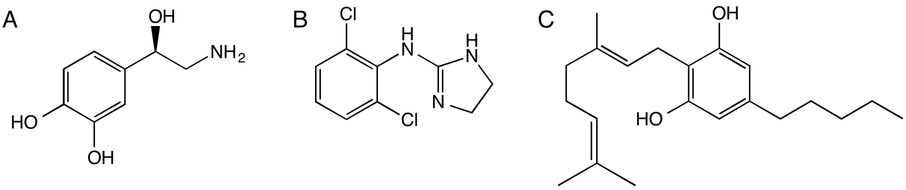


Figure 4.1.1: Structure representation of noradrenaline (panel A), clonidine (panel B) and CBG (panel c).

(α_2-ARs), which are in turn physiologically activated by noradrenaline (Figure 4.1.1, panel A).

This is an unexpected finding since the chemical structure of this cannabinoid is very different from all other known α_2-AR ligands, and no other cannabinoid has been reported to display such activity. All these observation opened up the possibility that CBG, like other established α_2-ARs agonists (i.e. clonidine, Figure 4.1.1, panel B), could show significant efficacy as anti-nociceptive agent when administered in vivo. Previous results obtained in collaboration with the Laboratory of Molecular Pharmacology of Prof. Barbara Costa at University of Milano-Bicocca suggest that CBG is an effective compound in reducing both inflammatory and neuropathic pain. Using antagonism studies we identified that the effects exerted against different type of pain are α_2-mediated, and specifically α_{2A}-subtype mediated.

To better clarify how CBG interacts with α_2-ARs and activate them, docking simulations were carried out using as target protein the murine α_{2A}-isoform. The protein model was build using homology modeling and molecular dynamic (MD) simulations were used to refine the model structure. Docking simulations have been then performed to identify CBG binding mode and suggest the key residues involved in ligand binding.

4.1.1.2 Methods

The primary aminoacidic sequences of murine α_{2A}-ARs were retrivied from the Uniprot database (Q01338). A search with BlastP (protein-protein Basic local alignment search tool) [186] in the PDB was then carried out to identify suitable template structures to use for homology modeling. Multiple alignments were performed using Clustal Omega program with defaults parameters, accessible through the European Bioinformatic Institute, in order to identify similarity among homologous sequences. PSIPRED,[187] JPRED [188] and DISOPRED [189] servers were used to predict the secondary structure content of the target sequences. MODELLER version 9.11 was used for homology modeling.[190] The quality of each model was assessed with different methods, as the DOPE score and the MOLPDF score (both from MODELLER output), VADAR [191] and AIDE.[192] The overall quality was estimated with respect to its geometry and energy (packaging defects, free energy of folding, core hydrophobic and charged residues).

The use of MD simulations performed using empirical force fields in explicit solvent could be very useful for the refinement of protein structures based on homology, depending on the quality of the force field. The models chosen have been used as initial structures for simulations. MD simulations were performed using GROMACS software package,[193] using AMBER99SB-ILDN force field.[194] The starting structure was soaked in a dodecahedral box of single point charge water molecules and simulated using periodic boundary conditions. The ionization state of charged residues was set to be consistent with neutral pH; in order to neutralize the overall charge of the system a number of water molecules equal to the protein net charge were replaced with Cl ions. The system was initially relaxed (steep descent, 10000 steps) and a short MD simulation of 50 ps was carried out to equilibrate the system at 300K (time step 1fs and thermal coupling constant of 0.006 ps). During thermalization, another short 50 ps MD simulation was performed. The system was finally pressurized through a series of three consecutive short 50 ps MD simulations, in order to slowly drive the simulation temperature and pressure to 300K and 1 bar, respectively. The pressure bath was coupled to the system every 0.1 ps and the pressure was reduced from 200 atm during the first simulation, to 100, and finally to 1 atm. Productive 20 ns MD simulations have been carried out in the thermal-isobaric (NPT) ensemble, using an external bath with a coupling constant of 0.1 ps at 300K. Helices residues were constrained during simulations by harmonical restraints since they are expected to be in the phospholipid bilayer and not in solution. The LINCS algorithm [195] was used to constrain bond lengths, allowing the use of 2 fs time step. Electrostatic interactions were calculated using the Particle-mesh Ewald summation scheme.

In order to evaluate the conformational variability among the structures sampled during the simulation was necessary build an RMSD matrix. After RMSD calculation

$$RMSD(t_1,t_2) = \sqrt{\left[\frac{1}{M}\sum_{i=1}^{N} m_i||r_i(t_1) - r_i(t_2)||^2\right]} \tag{4.1.5}$$

where $M = \sum_{I=1}^{N} m_i$ and r_i is the position of atom i at time t, results are ordered on a colour scale which ranges from blue to red, which respectively refer to RMSD values close to 0 (the structures are nearly identical) and high RMSD values, indicating structural divergence. The RMSD matrix is then processed to obtain structural clusters using the Gromos algorithm.[196] For each structure the number of other structures for which the RMSD is equal or less compared to a given cut-off is calculated; the structures found are called neighbour conformations. The structure with the highest number of neighbours is taken as the center of a cluster, and combined together with all its neighbours structure to form the first cluster. The structures of this cluster are thereafter eliminated from the pool of structures and the process is repeated until the pool of structures is empty. In this way, a series of non-overlapping clusters of structures was obtained. Cutoff used for cluster analysis of the trajectory was 0.225 nm.

Docking simulations were performed using DOCK 6.[181, 182, 197] Receptors and ligand atom types and charges were calculated using AM1-BCC [198] of antechamber package, available in the software UCSF Chimera,[199] in order to recognize atom and bond type, generating the residue topology file, finding force field parameters and supplying similar substitutes. Starting from receptor surface, spheres with a maximum radius of 4 Å and a minimum radius of 1.4 Å were generated. For every receptor, the largest sphere cluster represented the binding site and the receptor box delimiting was calculated using an additional 5Å boundary. Grid was calculated including all receptor atoms and the distance between grid point along each axes was of 0,3 Å. Non polar hydrogen were modeled using the all atom model: hydrogens attached to carbon have regular van der Waals well-depth and partial charge was not modified. During docking calculation a maximum number of 500 orientations were cycled. Docking optimization computed using AMBER score consisted in an initial minimization of 50 step, followed by a MD simulation of 1500 step and a further 50 step minimization. During the simulations only the ligand was allowed to move.

4.1.1.3 Results

Homology modelling allows to build a three dimensional model for a protein of unknown structure on the basis of sequence similarity to the template, a protein of known structure.[200, 201] The X-ray crystal structure of human dopamine D3 receptor was chosen as template (PDB code: 3PBL [202], 2.89 Å resolution) after an accurate research with BlastP. This receptor, like α_2-ARs, is a protein belonging to the class A GPCR and so shares many structural similarities. In addition dopamine D3 receptor X-ray structure is the only one with ELs (extracellular loops) solved. This characteristic is very important for our purpose since ligand binding occurs in the extracellular part of the protein, and a correct modeling of these structures is fundamental for an accurate docking study. Dopamine D3 receptor is also able to bind a catecholamine like α_2-ARs and we can suppose that its binding pocket shares many chemical and structural characteristics with ARs, which are able to bind catecholamine too. The multiple sequence alignment of α_{2A}-AR murine isoforms and template used to build the models is presented in Appendix. Since some long or extremely unstructured segments could give rise to artifacts in the following analysis, they were not modelled. Specifically, the intracellular loop 3 which is a disordered region of 159 aa was excluded from the modelling. Its lack should not influence our studies since the docking simulations focus on the upper part of the receptors and the extracellular loops, where the binding site is located, and they do not involve the intracellular loops. A population of 10 models was produced differing in the side chain conformations. In order to validate the models and find out eventual structural defects different approaches were used (see Methods section 4.1.1.2 and Appendix). According to geometry, energy and local structures one model was chosen (see figure 4.1.2).

To refine model structure a MD simulation was performed. The stability of MD trajectories was evaluated to verify that energetic and overall structural properties were stable during the whole simulation time. In particular, one of the structural parameters that require a longer time to converge is RMSD with respect to an initial reference structure. According to this parameter, the first 8 ns of the trajectory have been discarded from further analyses since the system was not equilibrated. To identify representative average structures from the remaining 12 ns MD trajectory, a structural clustering based on RMSD matrices was performed. The RMSD matrix (see Appendix) was built from

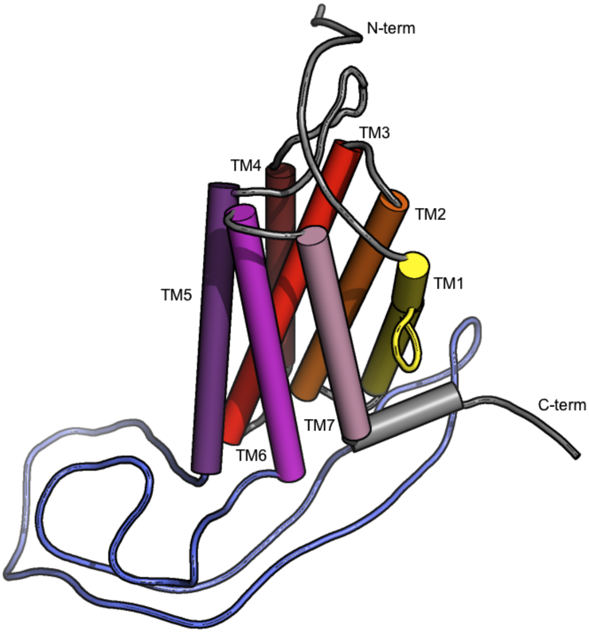


Figure 4.1.2: Model structure of murine α_{2A}-AR. Seven TM helices interconnected by six loop: IL1 between TM1 and TM2, EL1 between TM2 and TM3, IL2 between TM3 and TM4, EL2 between TM4 and TM5, IL3 between TM5 and TM6 and EL3 between TM6 and TM7.

	Noradrenaline	Clonidine	CBG
Structure 1	-17.4	-17.7	-25.5
Structure 2	-13.2	-11.9	-20.7
Structure 3	-19.5	-18.6	-26.2
Structure 4	-20.0	-21.2	-26.4
Structure 5	-21.4	-18.6	-27.1

Table 4.1: Docking results of NA, clonidine and CBG according to AMBER score (kcal mol^{-1})

the trajectory in order to evaluate the conformational variability among the sampled structures during the simulation. To cluster the conformations explored by the protein during the simulation, the RMSD matrix was processed using Gromos algorithm with a 0.225 nm cutoff and five different clusters were identified and sorted by decreasing numerosity of structures, so that cluster 1 is the most populated. No evident structural differences were identified among the clusters except for loops and side chain orientation. The five structure retrieved from cluster analysis have been used as starting point for the molecular docking simulations.

The binding site of α_{2A}-AR was deduced from mutagenesis data and the cavity containing it is funnel-shaped, with the upper and wider part containing the binding residues and a narrow neck in the direction of the intracellular space.[203] As told in Methods section 4.1.1.2, a cluster of spheres selected as binding site is needed to dock the ligand. The largest spheres cluster generated by DOCK 6 represented for all the five conformations the binding site. The run of flexible docking was performed using the five structure and noradrenaline, clonidine and CBG as ligands (the natural ligand, an already known nociceptive agent and the molecule to test respectively). The best results for every ligand according to AMBER score are shown in Table 4.1 and Figure 4.1.3.

Values of energy binding of noradrenaline and clonidine are very similar, a result supported also by previous sperimental work which identified comparable K_i values of these two compound.[204] CBG on the contrary shows an higher binding affinity. The receptor–ligand interactions were analyzed for each different conformation of α_{2A}-AR. The binding modes of the ligands we identified in the five conformations (structure 1, 2, 3, 4, 5, Table 4.1) are generally con-

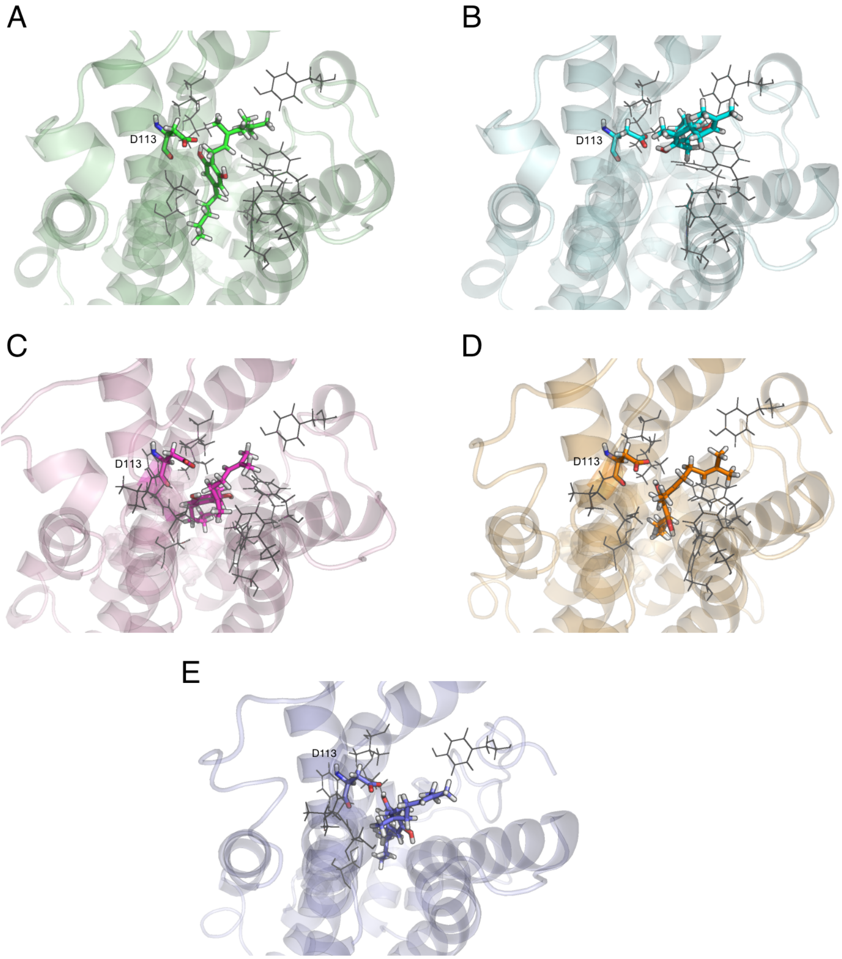


Figure 4.1.3: Docked conformations of CBG in: A, structure 1; B, structure 2; C, structure 3; D, structure 4; E, structure 5.

sistent with previous results. In addition to residues previously identified as fundamental in agonists binding in α_{2A}-ARs CBG can interact with many other aminoacids, especially hydrophobic ones. It's a generally accepted idea that α_2-AR activation involves transmembrane (TM) 3, 5 and 6 which can, through a conformational change, promote intracellular loop 3 movement and G protein activation.[205] The fundamental residues involved in this activation are Asp-113 in TM 3, Cys-201 and Cys-205 in TM5 and others Phe in TM6 and TM7. In phenetylamines the charged amine group is able to interact with Asp-113 with an hydrogen bond and the catecholic OH groups most probably interact with Cys-201, enabling TM3 and TM5 to move closer toward each other.[205] When CBG binds the receptor is able to interact with Asp-113 through one of the OH groups and, due to its steric encumbrance, it can interact with other aminoacids in the binding site. It can be supposed that the hydrophobic nature of both CBG and the binding site facilitate the binding, and the high number of interactions allow the conformational change of the receptor and the consequent activation. This mechanism of activation seems less specific compared to NA and clonidine one, but its explanation can lie once more in the chemical characteristic of CBG. While NA and clonidine have charged or polar functional groups that can easily orient the molecules in the binding site according to aminoacidic environment, CBG has only the two polar OH groups that can establish an hydrogen bond with Asp-113 (figure 4.8.1). In addiction CBG has two long symmetric aliphatic chain branching off the aromatic ring that can orient the molecule in the binding site without any specific interactions except hydrophobic ones; thus, as is possible to see in Figure 4.1.3, resulting CBG docking poses are slightly different within the five different receptor conformations. Despite of docking simulations, CBG seems to bind the receptor in different ways, but we can not be certain of this because it may be an artifact of the model. In any case it remains an anchor that Asp-113 play a fundamental role even in CBG binding and possibly receptor activation.

4.1.1.4 Conclusions

CBG appears to be a valid alternative to classic α_2-ARs agonists as it doesn't show collateral effects related to central activity and could be used in multidrug therapy for pain produced by multiple etiological factors. Our docking studies revealed that CBG is able to enter the binding pocket of the receptor and interact with the key residue Asp-113, known to be the trigger of the receptor activation. Since docking is still far from being considered an accurate or fully validated methodology, structural data from spectroscopies of the receptor will be necessary to understand if the different CBG binding poses identified are real existing. In addition, to gain a wide comprehension of CBG molecular mechanism, displacement binding assay with radio-ligands could be useful to obtain the values of K_d of the ligand to compare with affinity values obtained from molecular docking.

4.2 Glide

Glide algorithm 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 (i.e. $—CH_3$, $—NH_2$, $—NH_3{}^+$) 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 and membered 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 point 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

1 Å width. 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 and, 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.[183, 206] 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 [207] 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 Extra Precision GlideScore (XP GlideScore) scoring function.[208] The XP Glide scoring function is presented in equation 4.2.1

$$\text{XP GlideScore} = E_{coul} + E_{vdW} + E_{bind} + E_{penalty} \tag{4.2.1}$$

The main terms that favor binding are presented in equation 4.2.2, while those that hinder binding are presented in equation 4.2.3.

$$E_{bind} = E_{hyd\,enclosure} + E_{hb\,nn\,motif} + E_{hb\,cc\,motif} +$$

$$+E_{PI} + E_{hb\,pair} + E_{phobic\,pair} \tag{4.2.2}$$

$$E_{penalty} = E_{desolv} + E_{ligand\,strain} \tag{4.2.3}$$

where $E_{hyd\,enclosure}$ considers lipophilic ligand atoms enclosed on two sides by lipophilic protein atoms, $E_{hb\,nn\,motif}$ and $E_{hb\,cc\,motif}$ describe neutral-neutral and charged-charged hydrogen bonds respectively, E_{PI} takes into account pi and cation stacking interactions and $E_{hb\,pair}$ and $E_{phobic\,pair}$ are the terms for hydrogen bond and hydrophobic pairs respectively. In addition, for the penalty score, the ligand desolvation (E_{desolv}) and molecular strain ($E_{ligand\,strain}$) are taken into account.[208]

4.2.1 Synthetic sulfoglycolipids targeting the serine-threonine protein kinase Akt

4.2.1.1 Introduction

The serine-threonine protein kinase Akt, also known as protein kinase B, is a key component of the phosphoinositide 3-kinase (PI3K)-Akt-mTOR axis. Deregulated activation of this pathway is frequent in human tumors and Akt-dependent signaling appears to be critical in cell survival. PI3K activation generates 3-phosphorylated phosphatidylinositols that bind Akt pleckstrin homology (PH) domain. The blockage of Akt PH domain/phosphoinositides interaction represents a promising approach to interfere with the oncogenic potential of over-activated Akt. Several inhibitors targeting the PH domain have been proposed in the last years, including phosphoinositide analogues, alkylphospholipids and inositol phosphates.[209, 210, 211] Though synthetic carbohydrate-based mimics of inositols have already been reported,[212] a few examples on the use of glucose as scaffold for the design of Akt inhibitors have been described.[213, 214] Recently, the synthesis and the biological activity test of new sulphoglycolipids structurally related to the natural sulfoquinovosylacylglycerols (**1**, SQAG, Figure 4.2.1), a family of natural compounds associated with photosynthetic organisms in which sulfoquinovose (6-deoxy-6-sulfoglucose) is α-linked to the *sn*-3 position of an acylglycerol [213, 214] have been performed at the University of Milan-Bicocca and Milan respectively. Differently from SQAG, the studied compounds **2a-c** and **3a-c** (Figure 4.2.1) are based on a 2-O-β-D-glucopyranosyl-*sn*-glycerol scaffold, thus representing a novel class of potential Akt inhibitors acting as phosphoinositide mimics targeting the Akt PH domain. The structure of phosphatidylinositol-3-phosphate (PI3P, Figure 4.2.1), a natural ligand of Akt PH domain, as produced by PI3K, can resemble a suitably modified D-glucose scaffold, where specific functional groups can be introduced.

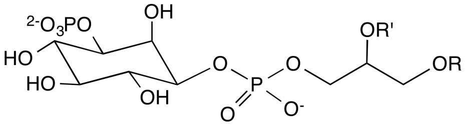
Thus, a series of D-glucopyranose compounds (Figure 4.2.1) endowed with the following key features needed for the interaction with the PH domain were synthesized: i) an anionic group mimicking the phosphate group in position 3 of the inositol (position 6 of the sugar), that should be able to form a salt bridge with a key arginine residue;[213] ii) a β-configured substituent at the anomeric carbon of D-glucopyranose (position 1 of inositol) possessing a hydrogen-donor heteroatom and a lipophilic group mimicking the diacyl glycerol moiety of phosphoinositides.

4.2.1.2 Methods

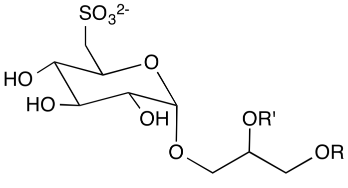
Docking studies were performed using the high resolution X-ray structure of PH domain of Protein kinase B/Akt (PDB entry: 1UNQ).[215] Ligands structures were built using the chemistry package Molecular Operating Environment (MOE) [Chemical Computing Group Inc. http://www.chemcomp.com/]. Docking simulations were performed using DOCK 6.[181, 182, 197] Receptors and ligand atom types and charges were calculated using AM1-BCC [198] of antechamber package, available in the software UCSF Chimera,[199] in order to recognize atom and bond type, generating the residue topology file, finding force field parameters and supplying similar substitutes. Starting from receptor surface, spheres with a maximum radius of 3 Å and a minimum radius of 1 Å were generated. The ensemble of spheres within 10Å of the bounded inositol-(1,3,4,5)-tetrakisphosphate (4IP) molecule was assumed to represent the binding site core, and box in which poses were generated was built ex-

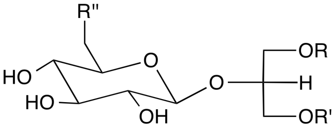
PI3P

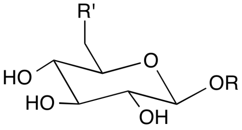
(R, R' = acyl chains)



1 (SQAG)







2a-c: R=R'=acyl, R''=SO_3^-

3a-c: R=acyl, R'=H, R''=SO_3^-

4: R=R'=H,R''=SO_3^-

5: R=R'=dodecanoyl, R''=OH

6: R=methyl, R'=SO_3^-

7: R=octyl, R'=SO_3^-

8: R=octyl, R'=NH_2

a: acyl = octadecanoyl b: acyl = dodecanoyl c: acyl = hexanoyl

Figure 4.2.1: Structure of D-gluco-pyranose-based glycoconjugates targeting Akt PH domain.

tending the binding site core by additional 20 Å along the Cartesian axes. Grid was calculated including all receptor atoms and the distance between grid point along each axes was of 0,3 Å. Non polar hydrogen were modeled using the all atom model: hydrogens attached to carbon have regular van der Waals well-depth and partial charge was not modified. During docking calculation a maximum number of 500 orientations were cycled. Docking optimization computed using AMBER score consisted in an initial minimization of 50 step, followed by a MD simulation of 1500 steps and a further 50 steps minimization. During the simulations only the ligand was allowed to move. When considering very flexible molecules, such as **2a**, several conformations (differing only for small differences in the orientation of the alkyl chains) clustered within a narrow energy range. However, for the sake of clarity, only the lowest energy poses are presented and discussed.

Glide docking was also performed.[183, 206, 208] The cartesian coordinates of the box were the same used for DOCK 6 docking. The force field used for docking is the OPLS-AA force field.[207]

4.2.1.3 Results

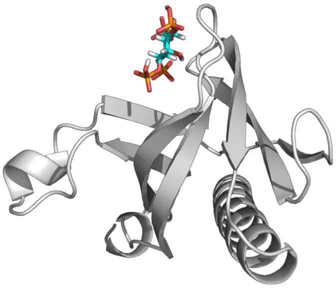
In an attempt to qualitatively define the mechanism by which the compounds tested in vitro could interact with Akt and inhibit its activity, docking simulations were carried out using the X-ray crystal structure of the PH domain of human Akt (PDB entry: 1UNQ,[215] 0.98 Å resolution). To assess the docking procedure, we performed the first run of docking using 4IP (inositol-(1,3,4,5)-tetrakisphosphate) as the ligand (in the crystal structure of Akt, the protein is co-crystallized with this organic cofactor). Since all the studied compounds are phosphatidylinositol (PIs) analogues, we used as docking space all the spheres generated by DOCK6 program within 10Å of the PIs binding site. The same conformational space was used to perform Glide docking. The resulting docking simulations poses are shown in Figure 4.2.2 (since DOCK 6 and Glide results are very similar only these latter are represented) and the relative docking energies in Table 4.2.

DOCK 6 and Glide results are in good agreement for both energy scores and docking poses. In addition, the interpretation of these results well fit with the experimental data obtained from ELISA tests performed in the laboratory of Pharmacology of Prof. Barbara Costa (Figure 4.2.3).

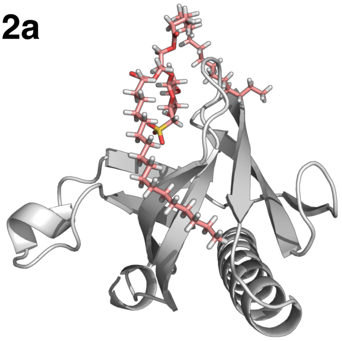
	Docking energies		Binding position of cycle compared to 4IP
	DOCK 6	Glide	
4IP / PI3P	-31.6 / -29.9	-26.4/-25.8	- / same
2a	-25.9	-23.2	same
2b	-19.2	-19.3	same
2c	-18.7	-17.2	same
3a	-23.5	-22.4	same
3b	-18.4	-17.2	same
3c	-15.8	-15.4	same
4	-6.5	-4.2	same
5	-18.5	-16.5	different
6	-19.3	-18.2	same
7	-20.4	-19.5	same
8	-4.5	-5.1	different

Table 4.2: Docking energies in kcal mol^{-1} of the tested compounds. In the column on the right is indicated if the binding position of the cycle of the compounds derived from PI3P is comparable to the one of the precursor.

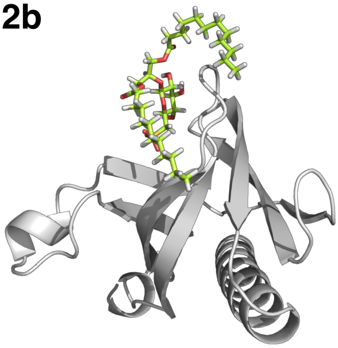
4IP



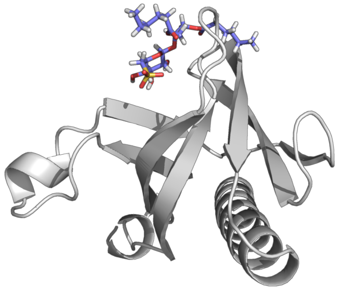
2a



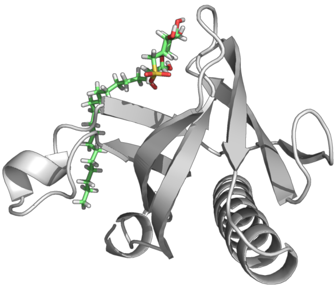
2b



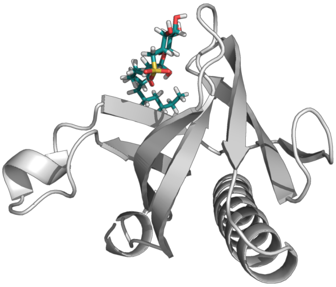
2c



3a



3b



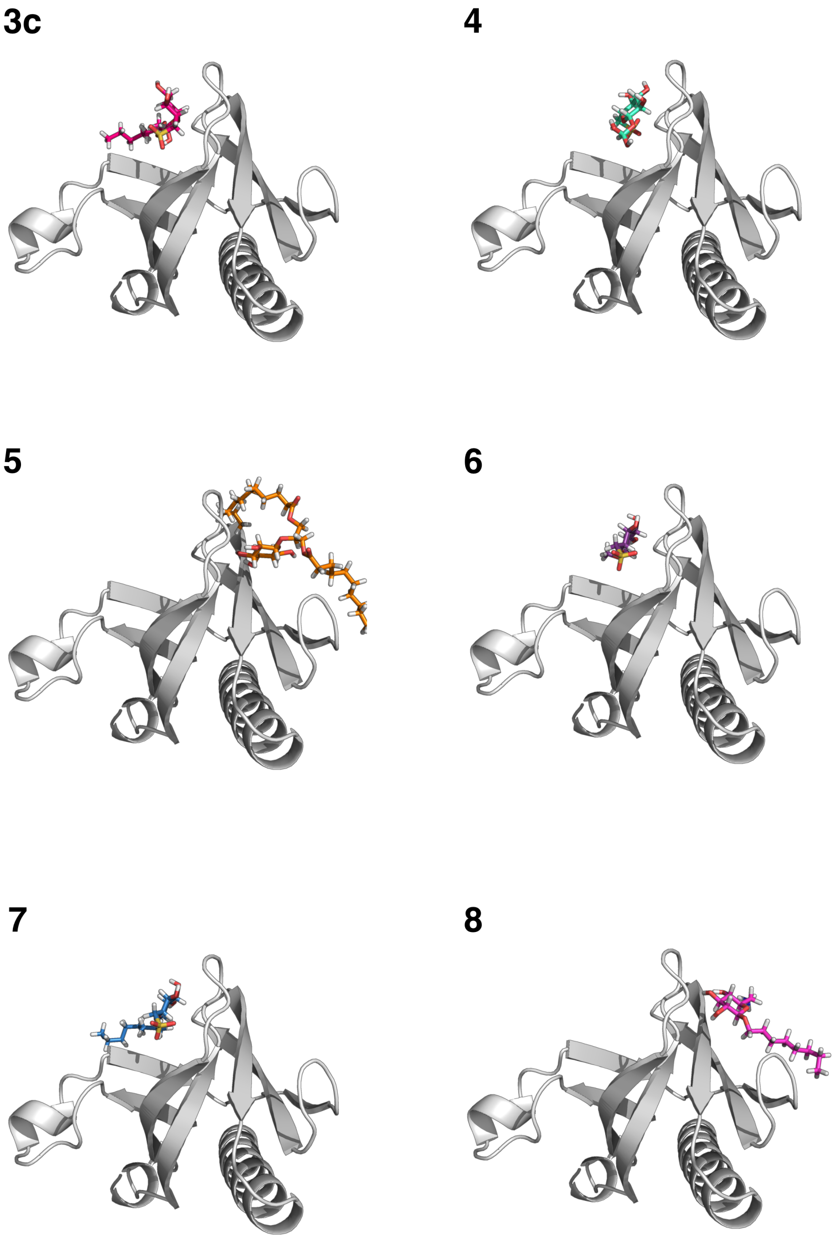


Figure 4.2.2: Docked conformations of 4IP (inositol-(1,3,4,5)-tetrakisphosphate) and **2**-**8** compounds.

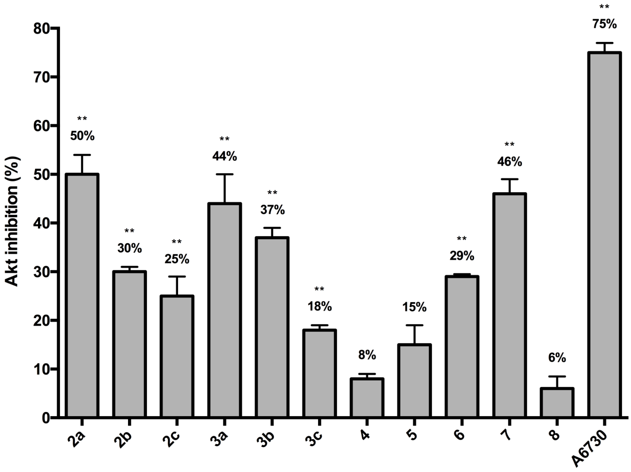


Figure 4.2.3: Inhibitory activity of compounds **2-8** against Akt. Akt activity was assayed through a specific ELISA test. Compounds were added at a concentration of 100 µM and the data represent the mean (±SD) of three independent experiments. * : $p < 0.05$, ** : $p < 0.01$ *vs* vehicle (Kruskal-Wallis non parametric ANOVA).

The docking of PI3P/4IP in the crystallographic structure ok Akt was used to validate the docking methodology. An RMSD of 0.2 Å between the crystallographic *vs* docked 4IP molecule proved the validity of the adopted docking methodology. The high inhibition capability of compounds **2a**, **3a**, and **7** (50%, 44% and 46 % respectively) is well supported by the high affinity calculated by docking simulations (a medium value between DOCK6 and Glide energies of 24.5, 22.9 and 20.0 kcal mol^{-1} respectively). The same can be observed for compounds with low inhibition activity on Akt (i.e. componds **4** and **8**).

A structural difference between compounds **5** and **8** and the rest of the series is the lack of the functional group SO_3^-. Since compounds **5** and **8**, when compared with other molecules, are found to bind in a different position, the sulfonate group seems to be fundamental for the correct binding of the molecule, and therefore for its biological activity. In particular the orientation of this group in a region of the protein surface rich in positively charged aminoacids (Lys-14, Arg-23, Arg-25, Figure 4.2.4) seems to stabilize the binding with an high number of hydrogen bonds. Only compound **4** showed a different orientation of the SO_3^- group, and this observation could also explain the weak docking score of this compound with Akt.

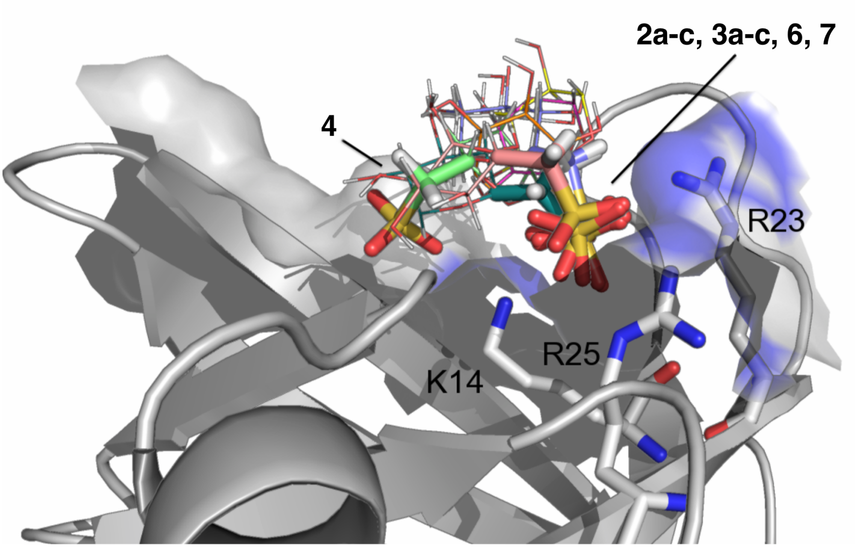


Figure 4.2.4: Focus on the orientation of the functional group SO_3^- in **2a-c**, **3a-c**, **4**, **6**, **7**. For every compound only group SO_3^- and the monosaccharide cycle are shown.

It must be underlined that also the length of the acyl chains plays a fundamental role in easing the binding energy of the compound. In fact, a longer chain appeared to provide a more stable interaction between Akt and the inhibitor than a shorter one.

4.2.1.4 Conclusions

In the present work, it has been proven the ability of synthetic sulfo-glycoconjugates derived from a D-glucosidic scaffold, designed to mimic phosphatidyl inositols, to impair Akt kinase activity. Docking studies and ELISA assay proved compounds **2a**, **3a** and **7** as the best inhibitors of Akt activity, and highlighted the role of sulfonate group for the binding. Further analysis are however required to assay their biological activity. In addition, the long lipophilic acyl chains of the natural PI3P substrate can be substituted in these compounds by shorter hydrophobic groups, since the long fatty acid chain is not needed for enzyme recognition, but only for membrane anchoring.[216]

Chapter 5

Evolution guided interactions in the immune system.

It is not the strongest of the species that survives,
nor the most intelligent that survives.
It is the one that is most adaptable to change.

Charles Darwin

Infections are thought to represent the major selective pressure for humans [217] and for all other living organisms. Encounters of hosts and pathogens result in the so-called *arms races*, whereby pathogens strive to develop countermeasures to evade host surveillance to achieve a successful infection and hosts are under pressure to evolve resistance to pathogens. Thus, when resistance and counter-resistance are at least partially genetically determined, cyclical adaptation and counteradaptation occur, and a genetic conflict is fuelled. The comparison among species gives a snapshot of selective events that happened over long timescales. Evolutionary pattern can in this way be identified by different approaches that compares genetic diversity and phylogenetic relationships.[218] Inter-species approaches to identify sequence substitutions or variations rely on the alignment of orthologous coding sequences and their site-by-site analysis. Among all possible substitutions, the analysis identify which substitutions are non-synonymous (aminoacid replacing) and which one are synonymous (non-aminoacid replacing). The observed number of non-synonymous differences for non-synonymous sites (dN) and the observed number of synonymous differences for synonymous sites (dS) are then estimated. Neutral evolution is a situation where the rate of aminoacid replacement accumulation is equal to the rate of silent changes. The ratio of dN and dS is so equal to 1. However, most of the aminoacid replacement are deleterious, and are so eliminated by selection. This situation is characterized by a majority of sites where dN/dS<1, a situation referred as *negative selection.* On the contrary, the positive pressure exerted by a pathogen may favour aminoacid replacements (i.e. changes that modify a crucial motif for host-pathogen interaction). Here, dN/dS ratio reaches a value greater than 1, a hallmark of *positive selection.* Therefore, evolutionary analysis can shed light on the sites in the host's and pathogen's proteins that are targeted by selection as a consequence of host-pathogen interaction.[218]

In this section will be illustrated a series of works resulting from the collaboration of the laboratory of Bioinformatics of dott.ssa Manuela Sironi at IRCCS E. Medea where different host-pathogen systems have been invastigated by evolutionary analysis. Structural biology has helped to understand and rationalize functional role of proteins, including their ability to interact and create complexes. Here, we used *in silico* mutagenesis, together with protein-protein docking and stability analysis to make a qualitative analysis of the results of the positive selection in pathogens and hosts proteins.

5.1 Protein-protein docking

One of the main challenges facing the proteomics community is the prediction of protein–protein interactions. The ultimate goal is to take the three-dimensional coordinates of two independently crystallized proteins which are known to interact, and to derive a model for the bound structure.[219, 220] Online it is possible to find a lot of programs and servers for protein-protein docking. In this manuscript we focused on the use of two different online servers for proteins docking: ClusPro and ROSIE. These programs have been successfully used for making high accuracy prediction in the CAPRI (Critical Assessment of PRedicted Interaction) challenge.[221, 222] ClusPro is an automated docking and discrimination method for the prediction of protein complexes.[223, 224] PIPER is an FFT-based docking program implemented in ClusPro server. It uses a pairwise interaction potential as part of its scoring function (see eq 5.1.1)

$$E = E_{attr} + w_1 E_{rep} + w_2 E_{elec} + w_3 E_{pair} \tag{5.1.1}$$

where E_{attr} and E_{rep} denote the attractive and repulsive contributions to the van der Waals interaction energy E_{vdw}, E_{elec} is an electrostatic energy term, and E_{pair} represents the desolvation contributions.[225] The coefficients *w1*, *w2*, and *w3* specify the weights of the corresponding terms, and are optimally selected for different types of docking problems. The second step of the algorithm is clustering the top 1000 structures generated using pairwise RMSD as the distance measure. For each docked conformation, the residues of the ligand that have any atom within 10 Å of any receptor atom are selected, and the RMSD for these residues from the same residues in all other 999 ligands is calculated. Based on the number of structures that a ligand has within a (default) cluster radius of 9 Å RMSD, the largest cluster is selected and ranked as center number 1. Then, the members of this cluster are removed, and the next largest cluster is selected and ranked. The biophysical meaning of clustering is isolating highly populated low energy basins of the energy landscape.[226] After clustering with this hierarchical approach, the ranked complexes are subjected to a straightforward (300 step and fixed backbone) van der Waals minimization using the CHARMM potential [227] to remove potential side chain clashes. The only refinement currently used in ClusPro is minimizing the Charmm energy of the structures generated by the docking.[227]

ROSIE on-line server implements RosettaDock, a multi-start, multi-scale Monte Carlo-based algorithm.[228] Initially, a low-resolution phase of the search is performed, including cycles of random rigid-body perturbations with a course-grained representation of side chains as single pseudo-atoms. Then, the high-resolution (all-atom, including hydrogens) phase of the search includes smaller rigid-body perturbations, with side-chain optimization via rotamer packing and continuous minimization,[229] and explicit gradient-based minimization of the rigid-body displacement. Scoring in the low-resolution phase includes residue-residue contacts and bumps. In the high-resolution phase, the energy is dominated by van der Waals energies,[228] orientation-dependent hydrogen bonding,[230] implicit Gaussian solvation,[231] side-chain rotamer probabilities [232] and a low-weighted electrostatics energy.[228] For a local docking perturbation run performed by the server, 1000 independent simulations are carried out to generate an ensemble of interacting proteins models .

5.1.1 T-cell regulation: the case of MIR2 and CD86

5.1.1.1 Introduction

T lymphocytes play a central role in the elicitation of effective immune responses and in the maintenance of immune homeostasis.[233] A pivotal role in T lymphocyte activation is played by the CD80 and CD86 costimulatory molecules. Although they represent only a subset of players in the complex process of T cell regulation, they have a central function in the elicitation of effective immune responses.[234] Consequently, several pathogens have evolved strategies that specifically target T cell regulatory molecules to facilitate the establishment of infection. In the case of CD86, several positions subject to diversifying selection were identified in the juxtamembrane (JM) and transmembrane (TM) regions, which interact with the MIR2 immunomodulator encoded by KSHV (Kaposi sarcoma-associated herpesvirus). [235] To deepen the understanding on how positive selection modulates MIR2/CD86 binding, we performed *ab initio* structure prediction and *in silico* docking analysis to study the interaction between these two proteins.

5.1.1.2 Methods

The sequences of human CD86 (P42081) and human herpes virus 8 MIR2 protein (P90489) were retrieved from Uniprot database. Secondary structure prediction was performed using PSIPRED.[187] Since no homologous proteins are available in the PDB, the structure of MIR2 (from Phe-85 to Arg-153) and CD86 (from Glu-235 to Lys-274) were obtained using QUARK,[236] a server for ab initio protein folding and protein structure prediction. CD86 variants were generated through MODELLER version 9.11, using as template the *ab initio* prediction. ROSIE (Rosetta Online Server that Includes Everyone) [237, 238] and ClusPro [224] were used for docking calculations. When using ROSIE, MIR2 and CD86 were positioned in order to allow the interaction between Asp-244 of CD86 and Ser-120 of MIR2.[235] CD86 variants at residues Val-260 and Trp-268 were generated by changing both aminoacids to alanine and by replacing them with aminoacids occurring in other mammalian species selected among those carrying an aspartic residue at position 244. As mentioned in the text, due to the relatively high number of positively selected sites, we also tried to model the JM region of CD86 (residues 235-248). However, the structure of this region is predicted to be random coil (a result confirmed by a PSIPRED analysis). Therefore, no docking analysis was possible.

5.1.1.3 Results

CD86 and MIR2 protein structures predictions were performed with *ab initio* techniques. The structure of the CD86 JM region (residues 235-248) was predicted to be random coil, preventing docking analysis; on the contrary, the TM region (249-274) was amenable to docking calculation. Protein-protein interaction predictions were performed with ROSIE and ClusPro on-line servers. The outputs of the two servers were comparable, with CD86 binding MIR2 in the same region and with a similar orientation. This revealed that the two CD86 positively selected sites, Val-260 and Trp-286, are located at the contact interface of the two proteins and can interact, mainly via hydrophobic contacts, with different residues of MIR2 (figure 5.1.1, panel A).

To verify the importance of the amino acids in positions 260 and 268, we created four different in silico CD86 variants (Val-260-Ala and Trp-268-Ala, or Val-260-Gly and Trp-268-Arg, these latter observed in some nonprimate mammals, figure 5.1.1, panel B): none of them was able to bind MIR2 with the same orientation as the human protein (figure 5.1.1, panel B in black). This analysis qualitatively confirms a crucial role of Val-260 and Trp-268 in the correct positioning of the two proteins during the interaction. The computed binding poses of the variant CD86 molecules with MIR2 are unlikely to occur in vivo, because of the presence of the phospholipid bilayer (which is not explicitly considered in the docking experiments). Therefore, these results suggest that variation at the positively selected sites renders interaction between CD86 and MIR2 unlikely or very weak.

5.1.1.4 Conclusions

The hypothesis that positive selection at T cell regulatory molecule genes is at least partially virus driven was directly tested for CD86. Indeed, the KSHV MIR2 ubiquitinase directly binds CD86 through its juxtramembrane (JM) and transmembrane (TM) regions.[235] Docking analysis indicated that the two positively selected sites in the TM region of CD86 are crucial for the interaction with MIR2. Therefore, the selective pressure exerted by MIR2 might

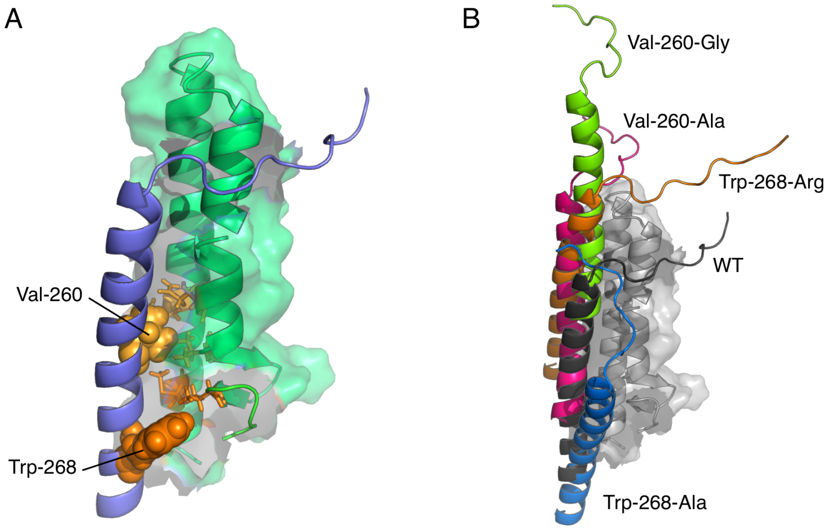


Figure 5.1.1: Structures of CD86 and MIR2. Panel A: docked structures of CD86 (blue) and MIR2 (green). The human CD86 amino acids Val-260 and Trp-268, which are involved in protein-protein interaction, are shown as yellow and orange spheres, respectively. The residues of MIR2 that interact with these two amino acids are represented as sticks. Panel B: docked structures of human CD86 (black) and four different variants: Val-260-Ala (fuchsia), Val-260-Gly (green), Trp-268-Ala (blue), and Trp-268-Arg (orange).

have driven the evolution of the CD86 TM region to decrease binding by viral-encoded ubiquitinases or to displace the ubiquitine ligase domain from its targets in the cytoplasmic domain of CD86.

5.2 Stability analysis

The availability of computational tools able to produce accurate estimations of the impact of aminoacid substitutions on protein stability is of crucial importance in a wide range of applications.[239] The rational design of modified proteins with controlled stability is of extreme importance in the biotechnological and environmental areas, where proteins are used for their catalytic or other functional activities. In addition, an improved understanding of the effect of naturally occurring disease-causing mutations on the molecular level is highly expected in the medical research area. Although multiple sequence alignments approaches still remain the prevalent tool in this context, stability changes prediction upon mutation have been approved as a relevant hint in the identification of disease-causing mutations.[240, 241] Different tools have been developed and are now available to study how mutations induce stability changes in proteins.

PoPMuSiC is a web server, that allows stability changes prediction resulting from point mutations in globular proteins.[242] The stability change resulting from a given point mutation in a protein is computed on the basis of the structure of the wild-type protein and a set of energy functions, which are used to estimate the folding free energy change upon mutation of a residue s_x into s_y, noted $\Delta\Delta\mathrm{G}_\mathrm{P}(s_x,s_y)$. More precisely, $\Delta\Delta\mathrm{G}_\mathrm{P}$ is expressed as a linear combination of different statistical potentials, two terms that depend on the volume of the wild-type and mutant amino acids, and an independent term. The statistical potentials are derived from a dataset of known protein structures and describe the correlations between various sequence or structure descriptors of the same aminoacids or of neighboring ones. The descriptors considered are, for each residue: the amino acid type, the torsion angles defining the backbone conformation and the solvent accessibility, and, for each pair of residues: the spatial distance between the average geometric centers of their side chains.[242] In addition, also the volume difference between the mutant and wild-type amino acid is taken into account.

FoldX is an empirical force field that aims to describe the energetic contributions to protein stability in simple empirical terms and has been implemented in the FoldX webserver.[243] A linear combination of empirical terms is used to calculate the protein free energy is showed in equation 5.2.1.

$$\Delta G = a\Delta G_{vdw} + b\Delta G_{solvH} + c\Delta G_{solvP} + d\Delta G_{wb} + e\Delta G_{hbond}$$

$$+ f\Delta G_{el} + g\Delta G_{kon} + hT\Delta S_{mc} + kT\Delta S_{sc} + k\Delta G_{clash} \tag{5.2.1}$$

The different terms considered include: interaction with the solvent in terms of desolvation separated into contributions from hydrophobic and polar groups (ΔG_{solvH} and ΔG_{solvP} respectively), van der Waals energy (ΔG_{vdw}) and hydrogen bond and electrostatic contributions (ΔG_{hbond} and ΔG_{el}). For protein complexes, an additional electrostatic contribution is calculated between atoms of different polypeptide chains (ΔG_{kon}). In FoldX the entropic penalty for fixing the backbone in a given conformation (ΔS_{mc}) is derived from a statistical analysis of the phi–psi distribution of a given amino acid as observed in a set of non-redundant high-resolution crystal structures. This entropy is scaled by the accessibility of the main chain atoms and energetics of hydrogen bond interactions made by the corresponding residue or its direct neighbours.[243]

I-Mutant 2.0 is a support vector machines (SVMs) based program that has been trained to predict both the direction (the ΔΔG sign) of the protein stability changes and the ΔΔG associated values. SVMs are learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. For the classification task and for assigning the ΔΔG values, basically two labels are identified: the increased protein stability ($\Delta\Delta G > 0$) and the destabilizing mutation ($\Delta\Delta G < 0$). To determine in which group the mutated protein has to be assigned, the temperature and the pH at which the stability of the mutated protein was experimentally determined, the mutation and the aminoacidic environment (spatially if the 3D structure is solved or the nearest sequence neighbors when only the protein sequence is available) are taken into account.[244]

5.2.1 Positive selection in ERAP2 has driven the recurrent appearance of protein-destabilizing variants

5.2.1.1 Introduction

Cell mediated immune responses are initiated by the recognition of an MHC/antigen complex on the surface of an antigen presenting cell by a T cell receptor. Whatever the nature of the presenting molecule, the MHC, the limited dimension of its cleft makes it impossible for macromolecules to be presented: only fragments deriving from the lysis of such molecules will be nested in the cleft. Most steps leading to the formation of MHC-peptide complexes have been defined. Peptides that will be embedded into the cleft of MHC molecules are initially processed by the proteasome, a complex structure located in the cytoplasm. Channels formed by TAP molecules allow peptides generated in the cytoplasm to be transported into the endoplasmic reticulum, where they may be trimmed at their N-terminal end by ERAP proteins. In this scenario, it is straightforward to imagine that both MHC and ERAP protein could be targeted by natural selection.[245]

Classic MHC molecules are encoded by genes that show extreme levels of polymorphism in most vertebrates and several studies have demonstrated that diversity at the peptide binding region is maintained by natural selection.[246] Thus, their role in adaptive immunity and their pattern of diversity indicate adaptation to a wide range of pathogen species leading to aminoacid diversification of the antigen binding cleft. The observation whereby several pathogens encode molecules that hijack specific components of the antigen processing and presentation pathway, like ERAP2, further supports this possibility.[247] The research group of Manuela Sironi identified three positively selected sites in ERAP2 (figure 5.2.1), which seem not be involved in proteolytic activity. Herein, we investigated the evolutionary history of ERAP2, trying to understand by stability analysis how positive selection has modified the functional activity of this protein.

5.2.1.2 Methods

Intra-protein interaction calculations were performed using PIC (Protein Interactions Calculator).[248] Stability analysis was carried out using three different methods. FoldX 3.0 [243] and PoPMuSiC (web-server version),[242] were used

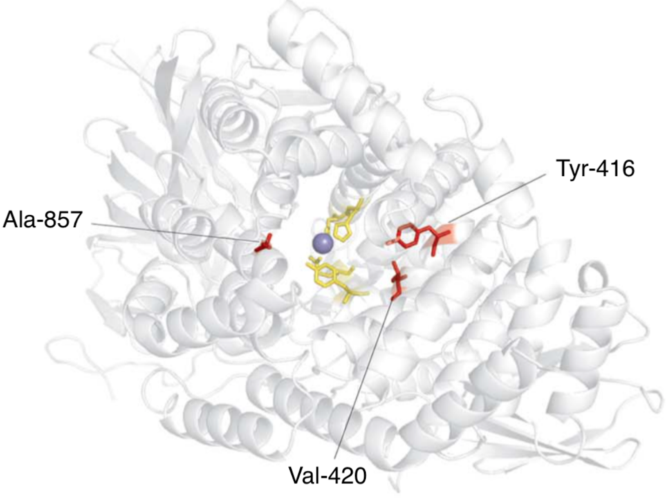


Figure 5.2.1: Ribbon diagram of ERAP2: positively selected sites are shown in red and those that coordinate the Zn ion (violet) in yellow

on the chain A of the X-ray structure of ERAP2 (PDB code: 3SE6). I-Mutant 2.0 [244] was used on the corresponding protein sequence retrieved from UniprotKB (Q6P179). In FoldX and I-Mutant the ΔΔG values are calculated as follows: $\Delta\Delta G = \Delta G_{mutant} - \Delta G_{wild-type}$. In FoldX and I-Mutant ΔΔG values > 0 indicate mutations that decrease protein stability, whereas in PoPMuSiC ΔΔG values > 0 are mark of mutation increasing protein stability. Therefore, PoPMuSiC ΔΔG values were multiplied by 21 to obtain homogeneous results. In the analysis carried out with FoldX 3D, the three-dimensional structure of the protein was repaired using the <RepairPDB> command. Mutations were introduced using the <BuildModel> command with <numberOfRuns> set to 5 and <VdWdesign> set to 0. Temperature (298K), ionic strength (0.05 M) and pH (7) were set to default values and the force-field predicted the water molecules on the protein surface. Residues His-370, His-374, Glu-393 and Tyr-455, which coordinates the zinc ion, were kept fixed during reparation and mutation procedures.

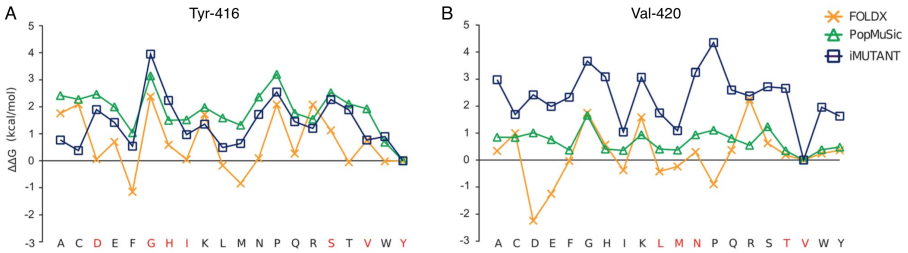


Figure 5.2.2: ΔΔG in kcal mol^{-1} for Tyr-416 (panel A), Val-420 (panel B) mutations to all other 19 residues of the ERAP2 structure or sequence; results are shown for FoldX, PopMuSiC, and I-Mutant.

5.2.1.3 Results

In ERAP2 three positively selected sites which do not seem to be involved in proteolytic activity have been identified. 3D-structure protein analysis indicated that the three residues are located on a helices shaping the internal cavity of the protein where the catalytic Zn ion is coordinated (Figure 5.2.1). Two of these residues are involved in several short-range interactions: Tyr-416 can interact hydrofobically with Leu-362, Phe-413, Trp-746, and Val-420 (and *vice-versa*). The same kind of interactions can be made by Val-420 with Phe-417; a side-chain side-chain H-bond can be formed by the OH group of Tyr-416 and the NH_2 group of Arg-366. Thus, we performed a stability analysis: Tyr-416 and Val-420 were mutated to all other residues through the use of three different methods. The tyrosine and valine at positions 416 and 420 are the most common aminoacids among the species included in the evolutionary analysis and represent the ancestral state residues (see Methods section). As shown in figure 5.2.2, the replacement of the two aminoacids led to changes of different magnitude in ΔG. Although the three programs yielded different ΔΔG values for every mutated residues, the trend was maintained (in particular between I-Mutant and PoPMuSiC programs) and indicated that replacement of Tyr-416 and Val-420 with any other aminoacid likely results in protein destabilization (i.e. positive ΔΔG values) (Figure 4C). These observations suggest that positive selection might have driven the recurrent appearance of destabilizing variants in ERAP2.

5.2.1.4 Conclusions

Natural selection might have acted at the ERAP2 locus to modulate protein stability and, consequently, abundance. Our stability analysis on the two positively selected sites Tyr-416 and Val-420 suggested that evolution has driven the appearance of destabilizing variants. Although the observation that protein-destabilizing variants have been favored during evolution might seem counter-intuitive, it should be noted that an ERAP2 that results in a truncated (and degraded) protein product is maintained in human populations.[249, 250] Also, some rodent species, including mice and rats, lack a functional ERAP2 gene, suggesting that loss or decreased abundance of ERAP2 protein might confer some advantage, possibly related to selective antigen trimming.[250]

Chapter 6

Conclusions and remarks

In this work we used computational methods to explore and disclose some phenomena at molecular level and they have proven to be powerful and effective tools to describe biotechnologically-relevant macromolecules. In addition to open up new questions on the systems studied, our results highlight how computational methods can support the analysis of chemical-biological systems at different levels.

Computational quantum chemistry has proven a valuable tool to gain a detailed description of noble-metal free bioinorganic systems for molecular H_2 production and CO_2 storage. The use of DFT in the study of [FeFe]-hydrogenase biomimetic models has allowed us to unveil some structural properties to achieve efficient synthetic models for both H_2 production and oxidation. Our findings highlight that the use of a rigid scaffolds covalently linked to cyanide-bound biomimetic diiron models like porphirin can lead to the stabilization of μ-CO geometries resembling the one disclosed by crystallographic studies of H_{red}. In addition, the use of a sufficiently electron poor diiron subcluster combined with an azadithiolate strap between the two iron atoms favours H_2 coordination and promote H_2 heterolytic cleavage. A more complex discussion has to be made for the proton reduction step, here discussed in subsection 2.1.2. First of all, the *in silico* study of redox potentials of diiron complexes can help identify the electronic properties that influence them on the molecular level.[84] With this purpose, to compute accurate theoretical redox potentials we had to tune the theoretical method on the experimental findings. Using a well studied [FeFe]-

hydrogenase biomimetic complex, we have tested which method provided the best match of theory *versus* experiment. The better performance of ΔE *versus* ΔG for redox potential calculation in Nernst equation ($\Delta G^{o}_{solv}{}^{, redox} = -FE^{o}$) that we found, can be ascribed to a cancelation error effect arising from approximations normally used to turn ΔE into ΔG. To overcome the "stable/less reactive µ-H vs unstable/reactive t-H" issue we searched for the combination of ligands that afforded species with $\Delta E°_{\text{t-H}-\mu\text{-H}} < 0$, thus indicating a higher (or at least as high as) redox reactivity of µ-H *versus* t-H. This point is relevant as µ-H are protonated isomers od diiron dithiolates that are easier to deal with (due to their higher stability *versus* t-H). Increasing the number of electron donors on the protonated [2FeH] core of [FeFe]-hydrogenases is predicted to alter $\Delta E°_{\text{t-H}-\mu\text{-H}}$ in a nonlinear way, up to a substitution extent at which subtle steric factors come into play entailing a surprising modulation of the redox gap. Variation upon reduction of the structural and energetic parameters shows that such a result is due to a steric strain (or destabilization), which affects differently the µ-H and t-H isomeric forms. The variation of the overall charge of the diiron dithiolates also tune the reduction potential, and more specifically, reducing the overall charge introducing nitrogen-containing linkers that are susceptible to being protonated by acids, tends to favor t-H reduction over µ-H. All these observations on [FeFe]-hydrogenase biomimetic models point out the necessity of incorporating different features on synthetic models to critically design systems of potential technological relevance, and highlight the essential role of computational modeling in supporting this process. DFT was a useful tool also to study the most efficient iron-based catalyst for CO_2 hydrogenation. We studied the free energy profiles associated with the intimate mechanism by which the CO_2-to-HCOOH reduction is performed. Here, the application of the *energetic span model* developed by Kozuch and Shaik [99, 100] proved to be crucial for the understanding of the molecular determinants characterizing the catalytic activity, nonetheless inclusion of thermodinamic criteria is relevant to fully rationalize the experimental scenario.

The use of TD-DFT with experimental spectroscopies has revealed a useful tool for the study of metal-boud transferrin. The presence of low glycosilated forms of this protein have been found to be the most specific and sensitive parameter for the diagnosis of chronic alcohol abuse. In this work we demonstrated that the Tf-Tb^{III} complex that could be used as a fluorescent label during the detection and we developed a computational method to study the metal binding site. Although in this case we have only preliminary results, the method we developed for the determination of the molecular model of the Tf interaction site with metal ions has proven to be promising.

Molecular modeling methods applied to medicinal chemistry has been of great value in the identification and development of novel promising compound. In collaboration with experimental groups of the University of Milan-Bicocca we conducted docking studies on promising compound for analgesic and anti-cancer purpose. The two programs used gave us comparable results, both from an energetical (docking energies) and a structural (docking poses) point of view. As Yu-Chian Chen suggests in his opinion paper *'Beware of docking!'* [175] docking analysis alone can create an inaccurate picture of ligand binding, in particular when used as a screening tool to identify hit compounds. The use of well-optimized systems, in combination with experimental results, can make docking a powerful tool to give a picture, an outline, on how a certain molecule is able to bind its molecular target and exert its effects. The same conclusions can be applied to protein-protein docking. Far to be the most-reliable method to describe inter-protein interactions, it can be a useful tool to make qualitative prediction on how protein binding is modified upon a mutation and explain how evolution drives interactions in the immune system.

Bibliography

[1] Richard H. Holm, Pierre Kennepohl, and Edward I. Solomon. Structural and Functional Aspects of Metal Sites in Biology. *Chemical Reviews*, 96(7):2239–2314, jan 1996.

[2] Margareta R. A. Blomberg and Per E. M. Siegbahn. A Quantum Chemical Approach to the Study of Reaction Mechanisms of Redox-Active Metalloenzymes. *The Journal of Physical Chemistry B*, 105(39):9375–9386, oct 2001.

[3] Douglas R. Hartree. 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(01):111, oct 1928.

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

[5] Erich Huckel. Quantentheoretische Beitrage zum Benzolproblem. *Zeitschrift fur Physik*, 70(3-4):204–286, mar 1931.

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

[7] Walter Kohn and Leu Jeu Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, 140(4A):A1133–A1138, nov 1965.

[8] Jerome A. Berson. Erich Hückel, Pioneer of Organic Quantum Chemistry: Reflections on Theory and Experiment. *Angewandte Chemie International Edition*, 35(2324):2750–2764, dec 1996.

[9] Warren J Hehre. *A Guide to Molecular Mechanics and Quantum Chemical Calculations.* 2003.

[10] Christopher J Cramer. *Essentials of Computational Chemistry Theories and Models*, volume 42. 2004.

[11] Stephen Wilson. *Chemistry by computer: an overview of the applications of computers in chemistry*, volume 1. 2005.

[12] Erich Runge and E. K. U. Gross. Density-Functional Theory for Time-Dependent Systems. *Physical Review Letters*, 52(12):997–1000, mar 1984.

[13] Robert van Leeuwen. Causality and Symmetry in Time-Dependent Density-Functional Theory. *Physical Review Letters*, 80(6):1280–1283, 1998.

[14] Andrew R. Leach. *Molecular Modelling: Principles and Applications.* 2001.

[15] Martin Karplus and J Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nature structural biology*, 9(9):646–52, sep 2002.

[16] Stefano Pascarella and Alessandro Paiardini. *Bioinformatica*. 2011.

[17] Paolo Avesani and Irwin King. *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes*, volume 1. 2014.

[18] Gerald D. Fasman. Prediction of Protein Structure and the Principles of Protein, 1989.

[19] Frank Jensen. *Introduction to Computational Chemistry*. John Wiley & Sons Ltd., Chichester, England, 2012.

[20] Susan Solomon, Gian-Kasper Plattner, Reto Knutti, and Pierre Friedlingstein. Irreversible climate change due to carbon dioxide emissions. *Proceedings of the National Academy of Sciences of the United States of America*, 106(6):1704–9, feb 2009.

[21] Wolfgang Lubitz and William Tumas. Hydrogen: An Overview. *Chemical Reviews*, 107(10):3900–3903, oct 2007.

[22] Cristina Gómez-Navarro, R. Thomas Weitz, Alexander M. Bittner, Matteo Scolari, Alf Mews, Marko Burghard, and Klaus Kern. Electronic Transport Properties of Individual Chemically Reduced Graphene Oxide Sheets. *Nano Letters*, 7(11):3499–3503, nov 2007.

[23] Pascale Chenevier, Laurent Mugherli, Sunita Darbe, Léa Darchy, Sylvain Di Manno, Phong D. Tran, Fabrice Valentino, Marina Iannello, Anne Volbeda, Christine Cavazza, and Vincent Artero. Hydrogenase enzymes: Application in biofuel cells and inspiration for the design of noble-metal free catalysts for H2 oxidation. *Comptes Rendus Chimie*, 16(5):491–505, 2013.

[24] Paulette M. Vignais and Bernard Billoud. Occurrence, classification, and biological function of hydrogenases: an overview. *Chemical Reviews*, 107(10):4206–72, oct 2007.

[25] Paulette M. Vignais, Bernard Billoud, and Jacques Meyer. Classification and phylogeny of hydrogenases. *FEMS Microbiology Reviews*, 25(4):455–501, aug 2001.

[26] Juan C. Fontecilla-Camps, Anne Volbeda, Christine Cavazza, and Yvain Nicolet. Structure/Function Relationships of [NiFe]- and [FeFe]-Hydrogenases. *Chemical Reviews*, 107(10):4273–4303, oct 2007.

[27] Juan C. Fontecilla-Camps, Patricia Amara, Christine Cavazza, Yvain Nicolet, and Anne Volbeda. Structure-function relationships of anaerobic gas-processing metalloenzymes. *Nature*, 460(7257):814–22, aug 2009.

[28] Brant K. J. Pohorelic, Johanna K. Voordouw, Elisabeth Lojou, Alain Dolla, Jens Harder, and Gerrit Voordouw. Effects of deletion of genes encoding Fe-only hydrogenase of Desulfovibrio vulgaris Hildenborough on hydrogen and lactate metabolism. *Journal of Bacteriology*, 184(3):679–86, feb 2002.

[29] Martin Winkler, Steffen Kawelke, and Thomas Happe. Light driven hydrogen production in protein based semi-artificial systems. *Bioresource Technology*, 102(18):8493–500, sep 2011.

[30] Michael W. W. Adams. The structure and mechanism of iron-hydrogenases. *Biochimica et Biophysica Acta*, 1020(2):115–45, nov 1990.

[31] David S. Horner, Burkhard Heil, Thomas Happe, and T Martin Embley. Iron hydrogenases-ancient enzymes in modern eukaryotes. *Trends in Biochemical Sciences*, 27(3):148–53, mar 2002.

[32] Jacques Meyer. [FeFe] hydrogenases and their evolution: a genomic perspective. *Cellular and Molecular Life Sciences*, 64(9):1063–84, may 2007.

[33] John W. Peters, William N. Lanzilotta, Brian J. Lemon, and Lance C. Seefeldt. X-ray Crystal Structure of the Fe-Only Hydrogenase (CpI) from Clostridium pasteurianum to 1.8 Angstrom Resolution. *Science*, 282(5395):1853–1858, dec 1998.

[34] Yvain Nicolet, Claudine Piras, Pierre Legrand, Claude E. Hatchikian, and Juan C. Fontecilla-Camps. Desulfovibrio desulfuricans iron hydrogenase: the structure shows unusual coordination to an active site Fe binuclear center. *Structure*, 7(1):13–23, jan 1999.

[35] Yvain Nicolet, Brian J. Lemon, Juan C. Fontecilla-Camps, and John W. Peters. A novel FeS cluster in Fe-only hydrogenases. *Trends in Biochemical Sciences*, 25(3):138–143, mar 2000.

[36] Sven T. Stripp, Gabrielle Goldet, Caterina Brandmayr, Oliver Sanganas, Kylie A. Vincent, Michael Haumann, Fraser A. Armstrong, and Thomas Happe. How oxygen attacks [FeFe] hydrogenases from photosynthetic organisms. *Proceedings of the National Academy of Sciences of the United States of America*, 106(41):17331–6, oct 2009.

[37] Martin Winkler, Julian Esselborn, and Thomas Happe. Molecular basis of [FeFe]-hydrogenase function: an insight into the complex interplay between protein and catalytic cofactor. *Biochimica et Biophysica Acta*, 1827(8-9):974–985, aug 2013.

[38] Anna Akhmanova, Frank Voncken, Theo van Alen, Angela van Hoek, Brigitte Boxma, Godfried Vogels, Marten Veenhuis, and Johannes H. P. Hackstein. A hydrogenosome with a genome. *Nature*, 396(6711):527–8, dec 1998.

[39] Yvain Nicolet, Antonio L. de Lacey, Xavier Vernède, Victor M. Fernandez, E. Claude Hatchikian, and Juan C. Fontecilla-Camps. Crystallographic and FTIR Spectroscopic Evidence of Changes in Fe Coordination Upon Reduction of the Active Site of the Fe-Only Hydrogenase from Desulfovibrio desulfuricans. *Journal of the American Chemical Society*, 123(8):1596–1601, feb 2001.

[40] Antonio J. Pierik, Marco Hulstein, Wilfred R. Hagen, and Simon P. J. Albracht. A low-spin iron with CN and CO as intrinsic ligands forms the core of the active site in [Fe]-hydrogenases. *European Journal of Biochemistry / FEBS*, 258(2):572–8, dec 1998.

[41] Alice S. Pereira, Pedro Tavares, Isabel Moura, José J. G. Moura, and Boi Hanh Huynh. Mossbauer characterization of the iron-sulfur clusters in desulfovibrio Vulgaris hydrogenase. *Journal of the American Chemical Society*, 123(12):2771–2782, mar 2001.

[42] Alexey Silakov, Brian Wenk, Eduard Reijerse, and Wolfgang Lubitz. 14 N HYSCORE investigation of the H-cluster of [FeFe] hydrogenase: evidence for a nitrogen in the dithiol bridge. *Physical Chemistry Chemical Physics*, 11(31):6592–6599, jul 2009.

[43] Carina E. Foster, Tobias Krämer, Annemarie F. Wait, Alison Parkin, David P. Jennings, Thomas Happe, John E. McGrady, and Fraser A. Armstrong. Inhibition of

[FeFe]-hydrogenases by formaldehyde and wider mechanistic implications for biohydrogen activation. *Journal of the American Chemical Society*, 134(17):7553–7, may 2012.

[44] Winfried Roseboom, Antonio L. De Lacey, Victor M. Fernandez, E Claude Hatchikian, and Simon P. J. Albracht. The active site of the [FeFe]-hydrogenase from Desulfovibrio desulfuricans. II. Redox properties, light sensitivity and CO-ligand exchange as observed by infrared spectroscopy. *Journal of Biological Inorganic Chemistry*, 11(1):102–18, jan 2006.

[45] Agnieszka Adamska, Alexey Silakov, Camilla Lambertz, Olaf Rüdiger, Thomas Happe, Edward Reijerse, and Wolfgang Lubitz. Identification and characterization of the "Super-Reduced" State of the H-Cluster in [FeFe] Hydrogenase: A New Building Block for the Catalytic Cycle? *Angewandte Chemie International Edition*, 51(46):11458–11462, nov 2012.

[46] Annemarie F Wait, Caterina Brandmayr, Sven T. Stripp, Christine Cavazza, Juan C. Fontecilla-Camps, Thomas Happe, and Fraser A. Armstrong. Formaldehyde–a rapid and reversible inhibitor of hydrogen production by [FeFe]-hydrogenases. *Journal of the American Chemical Society*, 133(5):1282–5, feb 2011.

[47] David W. Mulder, Eric M. Shepard, Jonathan E. Meuser, Neelambari Joshi, Paul W. King, Matthew C. Posewitz, Joan B. Broderick, and John W. Peters. Insights into [FeFe]-hydrogenase structure, mechanism, and maturation. *Structure*, 19(8):1038–52, aug 2011.

[48] Maurizio Bruschi, Claudio Greco, Markus Kaukonen, Piercarlo Fantucci, Ulf Ryde, and Luca De Gioia. Influence of the [2Fe]H Subcluster Environment on the Properties of Key Intermediates in the Catalytic Cycle of [FeFe] Hydrogenases: Hints for the Rational Design of Synthetic Catalysts. *Angewandte Chemie International Edition*, 48(19):3503–3506, 2009.

[49] Philipp Knörzer, Alexey Silakov, Carina E. Foster, Fraser A. Armstrong, Wolfgang Lubitz, and Thomas Happe. Importance of the Protein Framework for Catalytic Activity of [FeFe]-Hydrogenases. *Journal of Biological Chemistry*, 287(2):1489–1499, jan 2012.

[50] Jean-François Capon, Salah Ezzaher, Frédéric Gloaguen, François Y. Pétillon, Philippe Schollhammer, and Jean Talarmin. Electrochemical insights into the mechanisms of proton reduction by [Fe2(CO)6{mu-SCH2N(R)CH2S}] complexes related to the [2Fe](H) subsite of [FeFe]-hydrogenase. *Chemistry*, 14(6):1954–64, jan 2008.

[51] Frédéric Gloaguen and Thomas B. Rauchfuss. Small molecule mimics of hydrogenases: hydrides and redox. *Chemical Society Reviews*, 38(1):100–108, 2009.

[52] Cedric Tard and Christopher J. Pickett. Structural and Functional Analogues of the Active Sites of the [Fe]-, [NiFe]-, and [FeFe]-Hydrogenases. *Chemical Reviews*, 109(6):2245–2274, jun 2009.

[53] Marcetta Y. Darensbourg and Wolfgang Weigand. Hydrogenases. *European Journal of Inorganic Chemistry*, 2011(7):917–918, mar 2011.

[54] Wenguang Wang, Thomas B. Rauchfuss, Curtis E. Moore, Arnold L. Rheingold, Luca De Gioia, and Giuseppe Zampella. Crystallographic Characterization of a Fully Ro-

tated, Basic Diiron Dithiolate: Model for the Hred State? *Chemistry - A European Journal*, 19(46):15476–15479, nov 2013.

[55] Jesse W. Tye, Marcetta Y. Darensbourg, and Michael B. Hall. Correlation between computed gas-phase and experimentally determined solution-phase infrared spectra: models of the iron-iron hydrogenase enzyme active site. *Journal of Computational Chemistry*, 27(12):1454–62, sep 2006.

[56] Sabrina Munery, Jean-François Capon, Luca De Gioia, Catherine Elleouet, Claudio Greco, François Y. Pétillon, Philippe Schollhammer, Jean Talarmin, and Giuseppe Zampella. New FeI-FeI Complex Featuring a Rotated Conformation Related to the [2Fe]H Subsite of [FeâĂŞFe] Hydrogenase. *Chemistry - A European Journal*, 19(46):15458–15461, nov 2013.

[57] Filipp Furche, Reinhart Ahlrichs, Christof Hättig, Wim Klopper, Marek Sierka, and Florian Weigend. Turbomole. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(2):91–100, 2013.

[58] Karin Eichkorn, Florian Weigend, Oliver Treutler, and Reinhart Ahlrichs. Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. *Theoretical Chemistry Accounts*, 97(1-4):119–124, oct 1997.

[59] 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, apr 1994.

[60] John P. Perdew. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Physical Review B*, 33(12):8822–8824, jun 1986.

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

[62] Maurizio Bruschi, Claudio Greco, Piercarlo Fantucci, and Luca De Gioia. Structural and Electronic Properties of the [FeFe] Hydrogenase H-Cluster in Different Redox and Protonation States. A DFT Investigation. *Inorganic Chemistry*, 47(13):6056–6071, jul 2008.

[63] Giuseppe Zampella, Piercarlo Fantucci, and Luca De Gioia. Unveiling How Stereoelectronic Factors Affect Kinetics and Thermodynamics of Protonation Regiochemistry in [FeFe] Hydrogenase Synthetic Models: A DFT Investigation. *Journal of the American Chemical Society*, 131(31):10909–10917, aug 2009.

[64] Toshiko Miyake, Maurizio Bruschi, Ugo Cosentino, Carole Baffert, Vincent Fourmond, Christophe Léger, Giorgio Moro, Luca De Gioia, and Claudio Greco. Does the environment around the H-cluster allow coordination of the pendant amine to the catalytic iron center in [FeFe] hydrogenases? Answers from theory. *Journal of Biological Chemistry*, 18(6):693–700, aug 2013.

[65] Stefan Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of computational chemistry*, 27(15):1787–99, nov 2006.

[66] Mathieu Razavet, Stacey J. Borg, Simon J. George, Stephen P. Best, Shirley A. Fairhurst, and Christopher J. Pickett. Transient FTIR spectroelectrochemical and

stopped-flow detection of a mixed valence [Fe(I)-Fe(II)] bridging carbonyl intermediate with structural elements and spectroscopic characteristics of the di-iron sub-site of all-iron hydrogenase. *Chemical Communications*, (7):700–701, mar 2002.

[67] Daesung Chong, Irene P. Georgakaki, Rosario Mejia-Rodriguez, Jean Sanabria-Chinchilla, Manuel P. Soriaga, and Marcetta Y. Darensbourg. Electrocatalysis of hydrogen production by active site analogues of the iron hydrogenase enzyme: structure/function relationships. *Dalton Transactions*, (21):4158–4163, oct 2003.

[68] Frédéric Gloaguen, Joshua D. Lawrence, and Thomas B. Rauchfuss. Biomimetic Hydrogen Evolution Catalyzed by an Iron Carbonyl Thiolate. *Journal of the American Chemical Society*, 123(38):9476–9477, sep 2001.

[69] Xuan Zhao, Irene P. Georgakaki, Matthew L. Miller, Jason C. Yarbrough, and Marcetta Y. Darensbourg. H/D Exchange Reactions in Dinuclear Iron Thiolates as Activity Assay Models of Fe-H2ase. *Journal of the American Chemical Society*, 123(39):9710–9711, oct 2001.

[70] Frédéric Gloaguen, Joshua D. Lawrence, Thomas B. Rauchfuss, Marc Bénard, and Marie-Madeleine Rohmer. Bimetallic Carbonyl Thiolates as Functional Models for Fe-Only Hydrogenases. *Inorganic Chemistry*, 41(25):6573–6582, dec 2002.

[71] Stefanie Tschierlei, Sascha Ott, and Reiner Lomoth. Spectroscopically characterized intermediates of catalytic H2 formation by [FeFe] hydrogenase models. *Energy & Environmental Science*, 4(7):2340–2352, jul 2011.

[72] Giuseppe Zampella, Piercarlo Fantucci, and Luca De Gioia. DFT characterization of the reaction pathways for terminal- to μ-hydride isomerisation in synthetic models of the [FeFe]-hydrogenase active site. *Chemical Communications*, 46(46):8824–8826, nov 2009.

[73] Arndt R. Finkelmann, Martin T. Stiebritz, and Markus Reiher. Inaccessibility of the μ-hydride species in [FeFe] hydrogenases. *Chemical Science*, 5(1):215–221, nov 2013.

[74] Jarl Ivar van der Vlugt, Thomas B. Rauchfuss, C. Matthew Whaley, and Scott R. Wilson. Characterization of a Diferrous Terminal Hydride Mechanistically Relevant to the Fe-Only Hydrogenases. *Journal of the American Chemical Society*, 127(46):16012–16013, nov 2005.

[75] Salah Ezzaher, Jean-François Capon, Frédéric Gloaguen, François Y. Pétillon, Philippe Schollhammer, Jean Talarmin, Roger Pichon, and Nelly Kervarec. Evidence for the Formation of Terminal Hydrides by Protonation of an Asymmetric Iron Hydrogenase Active Site Mimic. *Inorganic Chemistry*, 46(9):3426–3428, apr 2007.

[76] Bryan E. Barton and Thomas B. Rauchfuss. Terminal Hydride in [FeFe]-Hydrogenase Model Has Lower Potential for H2 Production Than the Isomeric Bridging Hydride. *Inorganic Chemistry*, 47(7):2261–2263, apr 2008.

[77] Maria E. Carroll, Bryan E. Barton, Thomas B. Rauchfuss, and Patrick J. Carroll. Synthetic Models for the Active Site of the [FeFe]-Hydrogenase: Catalytic Proton Reduction and the Structure of the Doubly Protonated Intermediate. *Journal of the American Chemical Society*, 134(45):18843–18852, nov 2012.

[78] Aušra Jablonskyte, Joseph A. Wright, Shirley A. Fairhurst, Jamie N. T. Peck, Saad K. Ibrahim, Vasily S. Oganesyan, and Christopher J. Pickett. Paramagnetic Bridging Hydrides of Relevance to Catalytic Hydrogen Evolution at Metallosulfur Centers. *Journal of the American Chemical Society*, 133(46):18606–18609, nov 2011.

[79] Riccardo Zaffaroni, Thomas B. Rauchfuss, Danielle L. Gray, Luca De Gioia, and Giuseppe Zampella. Terminal vs Bridging Hydrides of Diiron Dithiolates: Protonation of Fe2(dithiolate)(CO)2(PMe3)4. *Journal of the American Chemical Society*, 134(46):19260–19269, nov 2012.

[80] James M. Camara and Thomas B. Rauchfuss. Mild redox complementation enables H2 activation by [FeFe]-hydrogenase models. *Journal of the American Chemical Society*, 133(21):8098–101, jun 2011.

[81] Wolfgang Lubitz, Hideaki Ogata, Olaf Rüdiger, and Edward Reijerse. Hydrogenases. *Chemical Reviews*, 114(8):4081–4148, apr 2014.

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

[83] Jian Li, Cindy L. Fisher, Jun L. Chen, Donald Bashford, and Louis Noodleman. Calculation of Redox Potentials and pKa Values of Hydrated Transition Metal Cations by a Combined Density Functional and Continuum Dielectric Theory. *Inorganic Chemistry*, 35(16):4694–4702, jan 1996.

[84] Lindsay E. Roy, Enrique R. Batista, and P. Jeffrey Hay. Theoretical Studies on the Redox Potentials of Fe Dinuclear Complexes as Models for Hydrogenase. *Inorganic Chemistry*, 47(20):9228–9237, oct 2008.

[85] 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.

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

[87] Li-Cheng Song, Zhi-Yong Yang, Hong-Zhu Bian, and Qing-Mei Hu. Novel Single and Double Diiron Oxadithiolates as Models for the Active Site of [Fe]-Only Hydrogenases. *Organometallics*, 23(13):3082–3084, jun 2004.

[88] Zhen Wang, Jian-Hui Liu, Cheng-Jiang He, Shi Jiang, Björn Åkermark, and Li-Cheng Sun. Azadithiolates cofactor of the iron-only hydrogenase and its PR3-monosubstituted derivatives: Synthesis, structure, electrochemistry and protonation. *Journal of Organometallic Chemistry*, 692(24):5501–5507, nov 2007.

[89] Chadwick A. Tolman. Steric effects of phosphorus ligands in organometallic chemistry and homogeneous catalysis. *Chemical Reviews*, 77(3):313–348, jun 1977.

[90] Joseph Chatt, C. T. Kan, G. Jeffery Leigh, Christopher J. Pickett, and David R. Stanley. Transition-metal binding sites and ligand parameters. *Journal of the Chemical Society, Dalton Transactions*, (10):2032–2038, jan 1980.

[91] Christopher J. Pickett. G. Wilkinson, R.D. Gillard, J.A. McCleverty (Eds.), Comprehensive Coordination Chemistry, vol. 1. Pergamon Press, Oxford, 1987.

[92] Alfred B. P. Lever. Electrochemical parametrization of metal complex redox potentials, using the ruthenium(III)/ruthenium(II) couple to generate a ligand electrochemical series. *Inorganic Chemistry*, 29(6):1271–1285, mar 1990.

[93] William A. Henderson and C. A. Streuli. The Basicity of Phosphines. *Journal of the American Chemical Society*, 82(22):5791–5794, nov 1960.

[94] Seigo Shima, Oliver Pilak, Sonja Vogt, Michael Schick, Marco S. Stagni, Wolfram Meyer-Klaucke, Eberhard Warkentin, Rudolf K. Thauer, and Ulrich Ermler. The Crystal Structure of [Fe]-Hydrogenase Reveals the Geometry of the Active Site. *Science*, 321(5888):572–575, jul 2008.

[95] Inigo Aguirre de Carcer, Antonio Di Pasquale, Arnold L. Rheingold, and D. Michael Heinekey. Active-Site Models for Iron Hydrogenases: Reduction Chemistry of Dinuclear Iron Complexes. *Inorganic Chemistry*, 45(20):8000–8002, oct 2006.

[96] Jean-François Capon, Saïd El Hassnaoui, Frédéric Gloaguen, Philippe Schollhammer, and Jean Talarmin. N-Heterocyclic Carbene Ligands as Cyanide Mimics in Diiron Models of the All-Iron Hydrogenase Active Site. *Organometallics*, 24(9):2020–2022, apr 2005.

[97] Gustav Berggren, Agnieszka Adamska, Camilla Lambertz, Trevor R. Simmons, Julian Esselborn, Mohamed Atta, Serge Gambarelli, Jean-Marie Mouesca, Edward J. Reijerse, Wolfgang Lubitz, Thomas Happe, Vincent Artero, and Marc Fontecave. Biomimetic assembly and activation of [FeFe]-hydrogenases. *Nature*, 499(7456):66–69, jul 2013.

[98] Joseph A. Wright and Christopher J. Pickett. Towards a Functional Model of the [FeFe]-Hydrogenase: Dihydrogen Oxidation. *ChemCatChem*, 4(11):1723–1724, nov 2012.

[99] Sebastian Kozuch and Sason Shaik. A combined kinetic-quantum mechanical model for assessment of catalytic cycles: application to cross-coupling and Heck reactions. *Journal of the American Chemical Society*, 128(10):3355–65, mar 2006.

[100] Sebastian Kozuch and Sason Shaik. How to conceptualize catalytic cycles? The energetic span model. *Accounts of chemical research*, 44(2):101–10, feb 2011.

[101] 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, jan 1988.

[102] Axel D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, apr 1993.

[103] Tianbiao Liu, Xiaoping Wang, Christina Hoffmann, Daniel L. DuBois, and R. Morris Bullock. Heterolytic Cleavage of Hydrogen by an Iron Hydrogenase Model: An Fe-H - H-N Dihydrogen Bond Characterized by Neutron Diffraction. *Angewandte Chemie International Edition*, 53(21):5300–5304, may 2014.

[104] Shunichi Fukuzumi. Bioinspired Energy Conversion Systems for Hydrogen Production and Storage. *European Journal of Inorganic Chemistry*, 2008(9):1351–1362, mar 2008.

[105] Philip G. Jessop, Ferenc Joó, and Chih-Cheng Tai. Recent advances in the homogeneous hydrogenation of carbon dioxide. *Coordination Chemistry Reviews*, 248(21-24):2425–2442, dec 2004.

[106] Amanda J. Morris, Gerald J. Meyer, and Etsuko Fujita. Molecular approaches to the photocatalytic reduction of carbon dioxide for solar fuels. *Accounts of chemical research*, 42(12):1983–94, dec 2009.

[107] Mark Z. Jacobson. Review of solutions to global warming, air pollution, and energy security. *Energy & Environmental Science*, 2(2):148, jan 2009.

[108] John A. Turner. A Realizable Renewable Energy Future. *Science*, 285(5428):687–689, jul 1999.

[109] John A Turner. Sustainable hydrogen production. *Science*, 305(5686):972–4, aug 2004.

[110] Louis Schlapbach and Andreas Züttel. Hydrogen-storage materials for mobile applications. *Nature*, 414(6861):353–8, nov 2001.

[111] Yoshio Inoue, Hitoshi Izumida, Yoshiyuki Sasaki, and Harukichi Hashimoto. Catalytic fixation of carbon dioxide to formic acid by transition-metal complexes under mild conditions. *Chemistry Letters*, (8):863–864, mar 1976.

[112] Philip G. Jessop, Takao Ikariya, and Ryoji Noyori. Homogeneous catalytic hydrogenation of supercritical carbon dioxide. *Nature*, 368(6468):231–233, mar 1994.

[113] Yuichiro Himeda. Conversion of CO2 into Formate by Homogeneously Catalyzed Hydrogenation in Water: Tuning Catalytic Activity and Water Solubility through the Acid-Base Equilibrium of the Ligand. *European Journal of Inorganic Chemistry*, 2007(25):3927–3941, sep 2007.

[114] Christopher Federsel, Albert Boddien, Ralf Jackstell, Reiko Jennerjahn, Paul J. Dyson, Rosario Scopelliti, Gabor Laurenczy, and Matthias Beller. A Well-Defined Iron Catalyst for the Reduction of Bicarbonates and Carbon Dioxide to Formates, Alkyl Formates, and Formamides. *Angewandte Chemie International Edition*, 49(50):9777–9780, 2010.

[115] Cynthia Rice, Soomin Ha, Richard I. Masel, Piotr Waszczuk, Andrzej Wieckowski, and Tom Barnard. Direct formic acid fuel cells. *Journal of Power Sources*, 111(1):83–89, sep 2002.

[116] Horst Schutte, Josef Flossdorf, Hermann Sahm, and Maria-Regina Kula. Purification and Properties of Formaldehyde Dehydrogenase and Formate Dehydrogenase from Candida boidinii. *European Journal of Biochemistry*, 62(1):151–160, feb 1976.

[117] Apostolos Alissandratos, Hye-Kyung Kim, Hayden Matthews, James E Hennessy, Amy Philbrook, and Christopher J Easton. Clostridium carboxidivorans strain P7T recombinant formate dehydrogenase catalyzes reduction of CO(2) to formate. *Applied and environmental microbiology*, 79(2):741–4, jan 2013.

[118] Shuning Wang, Haiyan Huang, Jörg Kahnt, and Rudolf K. Thauer. A reversible electron-bifurcating ferredoxin- and NAD-dependent [FeFe]-hydrogenase (HydABC) in Moorella thermoacetica. *Journal of Bacteriology*, 195(6):1267–75, mar 2013.

[119] Shuning Wang, Haiyan Huang, Jörg Kahnt, Alexander P. Mueller, Michael Köpke, and Rudolf K. Thauer. NADP-specific electron-bifurcating [FeFe]-hydrogenase in a functional complex with formate dehydrogenase in Clostridium autoethanogenum grown on CO. *Journal of Bacteriology*, 195(19):4373–86, oct 2013.

[120] Jiafu Shi, Yanjun Jiang, Zhongyi Jiang, Xueyan Wang, Xiaoli Wang, Shaohua Zhang, Pingping Han, and Chen Yang. Enzymatic conversion of carbon dioxide. *Chemical Society Reviews*, 44(17):5981–6000, oct 2015.

[121] Mark W. Farlow and Homer Adkins. The Hydrogenation of Carbon Dioxide and a Correction of the Reported Synthesis of Urethans. *Journal of the American Chemical Society*, 57(11):2222–2223, nov 1935.

[122] Walter Leitner. Carbon Dioxide as a Raw Material: The Synthesis of Formic Acid and Its Derivatives from CO2. *Angewandte Chemie International Edition*, 34(20):2207–2221, nov 1995.

[123] Karl W. Frese. *Electrochemical and Electrocatalytic Reactions of Carbon Dioxide*. Elsevier, 1993.

[124] Raimo Tenhunen, Harvey S. Marver, and Rudi Schmid. The enzymatic conversion of heme to bilirubin by microsomal heme oxygenase. *Proceedings of the National Academy of Sciences of the United States of America*, 61(2):748–755, 1968.

[125] Ryo Tanaka, Makoto Yamashita, and Kyoko Nozaki. Catalytic hydrogenation of carbon dioxide using Ir(III)-pincer complexes. *Journal of the American Chemical Society*, 131(40):14168–9, oct 2009.

[126] Elisabeth Graf and Walter Leitner. Direct formation of formic acid from carbon dioxide and dihydrogen using the [[Rh(cod)Cl]2]?Ph2P(CH2)4PPh2 catalyst system. *Journal of the Chemical Society, Chemical Communications*, (8):623, jan 1992.

[127] Franz Gassner and Walter Leitner. Hydrogenation of carbon dioxide to formic acid using water-soluble rhodium catalyststs. *Journal of the Chemical Society, Chemical Communications*, (19):1465, jan 1993.

[128] Philip G. Jessop, Yi Hsiao, Takao Ikariya, and Ryoji Noyori. Homogeneous Catalysis in Supercritical Fluids: Hydrogenation of Supercritical Carbon Dioxide to Formic Acid, Alkyl Formates, and Formamides. *Journal of the American Chemical Society*, 118(2):344–355, jan 1996.

[129] Atsushi Urakawa, Fabian Jutz, Gábor Laurenczy, and Alfons Baiker. Carbon dioxide hydrogenation catalyzed by a ruthenium dihydride: a DFT and high-pressure spectroscopic investigation. *Chemistry*, 13(14):3886–99, jan 2007.

[130] Christopher Federsel, Ralf Jackstell, and Matthias Beller. State-of-the-Art Catalysts for Hydrogenation of Carbon Dioxide. *Angewandte Chemie International Edition*, 49(36):6254–6257, 2010.

[131] Robert Langer, Gregory Leitus, Yehoshoa Ben-David, and David Milstein. Efficient hydrogenation of ketones catalyzed by an iron pincer complex. *Angewandte Chemie International Edition*, 50(9):2120–4, feb 2011.

[132] Yoshiyuki Sasaki, Yoshio Inoue, and Harukichi Hashimoto. Reaction of carbon dioxide with butadiene catalysed by palladium complexes. Synthesis of 2-ethylidenehept-5-en-4-olide. *Chemical Communications*, (15):605, jan 1976.

[133] G.O. Evans and C.J. Newell. Conversion of CO2, H2, and alcohols into formate esters using anionic iron carbonyl hydrides. *Inorganica Chimica Acta*, 31:L387–L389, jan 1978.

[134] Chih-Cheng Tai, Tangel Chang, Brentin Roller, and Philip G. Jessop. High-pressure combinatorial screening of homogeneous catalysts: hydrogenation of carbon dioxide. *Inorganic chemistry*, 42(23):7340–1, nov 2003.

[135] Carolin Ziebart, Christopher Federsel, Pazhamalai Anbarasan, Ralf Jackstell, Wolfgang Baumann, Anke Spannenberg, and Matthias Beller. Well-Defined Iron Catalyst for Improved Hydrogenation of Carbon Dioxide and Bicarbonate. *Journal of the American Chemical Society*, 134(51):20701–20704, dec 2012.

[136] Albert Boddien, Dörthe Mellmann, Felix Gärtner, Ralf Jackstell, Henrik Junge, Paul J. Dyson, Gábor Laurenczy, Ralf Ludwig, and Matthias Beller. Efficient Dehydrogenation of Formic Acid Using an Iron Catalyst. *Science*, 333(6050):1733–1736, sep 2011.

[137] Claudio Bianchini, Maurizio Peruzzini, and Fabrizio Zanobini. An exceptionally stable cis-(hydride)(η2-dihydrogen) complex of iron. *Journal of Organometallic Chemistry*, 354(2):C19–C22, oct 1988.

[138] Claudio Bianchini, Franco Laschi, Maurizio Peruzzini, Francesca M. Ottaviani, Alberto Vacca, and Piero Zanello. Electrochemistry as a diagnostic tool to discriminate between classical M-(H)2 and nonclassical M-(H2) structures within a family of dihydride and dihydrogen metal complexes. *Inorganic Chemistry*, 29(18):3394–3402, sep 1990.

[139] Claudio Bianchini, Dante Masi, Maurizio Peruzzini, Maurizio Casarin, Chiara Maccato, and Gian Andrea Rizzi. Ab Initio and Experimental Studies on the Structure and Relative Stability of the cis-Hydride-η2-Dihydrogen Complexes [{P(CH2CH2PPh2)3}M(H)(η2-H2)]+ (M = Fe, Ru). *Inorganic Chemistry*, 36(6):1061–1069, mar 1997.

[140] Xinzheng Yang. Mechanistic insights into iron catalyzed dehydrogenation of formic acid: β-hydride elimination vs. direct hydride transfer. *Dalton Transactions*, 42(33):11987–11991, aug 2013.

[141] Greetje de Jong, Jitse P. van Dijk, and H. G. van Eijk. The biology of transferrin. *Clinica Chimica Acta*, 190(1-2):1–46, sep 1990.

[142] J. Fletcher and E. R. Huehns. Function of Transferrin. *Nature*, 218(5148):1211–1214, jun 1968.

[143] Ching-Ming M. Chung. Stucture and function of transferrin. *Biochemical Education*, 12(4):146–154, oct 1984.

[144] Philip Aisen, Adela Leibman, and Jay Zweier. Stoichiometric and site characteristics of the binding of iron to human transferrin. *Journal of Biological Chemistry*, 253(6):1930–7, mar 1978.

[145] Helena Stibler. Carbohydrate-deficient transferrin in serum: a new marker of potentially harmful alcohol consumption reviewed. *Clinical chemistry*, 37(12):2029–37, dec 1991.

[146] Federica Bortolotti, Giorgia De Paoli, and Franco Tagliaro. Carbohydrate-deficient transferrin (CDT) as a marker of alcohol abuse: a critical review of the literature 2001-2005. *Journal of Chromatography B*, 841(1-2):96–109, sep 2006.

[147] H. G. van Eijk and Greetje de Jong. The physiology of iron, transferrin, and ferritin. *Biological Trace Element Research*, 35(1):13–24, oct 1992.

[148] Federica Bortolotti, Maristella Trettene, Rossella Gottardo, Marzia Bernini, Maria C. Ricossa, and Franco Tagliaro. Carbohydrate-deficient transferrin (CDT): a reliable indicator of the risk of driving under the influence of alcohol when determined by capillary electrophoresis. *Forensic Science International*, 170(2-3):175–8, aug 2007.

[149] Helena Stibler and K. G. Kjellin. Isoelectric focusing and electrophoresis of the CSF proteins in tremor of different origins. *Journal of the neurological sciences*, 30(2-3):269–85, dec 1976.

[150] Helena Stibler and Stefan Borg. Glycoprotein glycosyltransferase activities in serum in alcohol-abusing patients and healthy controls. *Scandinavian Journal of Clinical and Laboratory Investigation*, 51(1):43–51, feb 1991.

[151] Jason P Connor, Paul S Haber, and Wayne D Hall. Alcohol use disorders. *The Lancet*, sep 2015.

[152] Thomas M Maenhout, Guido Baten, Marc L. De Buyzere, and Joris R. Delanghe. Carbohydrate deficient transferrin in a driver's license regranting program. *Alcohol and Alcoholism*, 47(3):253–60, jan 2012.

[153] Eduardo Osuna, Maria D. Pérez-Cárceles, Miguel Moreno, Andrés Bedate, Josefa Conejero, José M. Abenza, Pedro Martínez, and Aurelio Luna. Vitreous humor carbohydrate-deficient transferrin concentrations in the postmortem diagnosis of alcoholism. *Forensic Science International*, 108(3):205–13, feb 2000.

[154] Torsten Arndt. Carbohydrate-deficient Transferrin as a Marker of Chronic Alcohol Abuse: A Critical Review of Preanalysis, Analysis, and Interpretation. *Clinical Chemistry*, 47(1):13–27, jan 2001.

[155] Joris R. Delanghe and Marc L. De Buyzere. Carbohydrate deficient transferrin and forensic medicine. *Clinica Chimica Acta*, 406(1-2):1–7, aug 2009.

[156] Michel Reynaud, Florence Hourcade, Francois Planche, Eliane Albuisson, Marie N. Meunier, and Roger Planche. Usefulness of carbohydrate-deficient transferrin in alcoholic patients with normal gamma-glutamyltranspeptidase. *Alcoholism, Clinical and Experimental Research*, 22(3):615–8, may 1998.

[157] Raye Z. Litten, John P. Allen, and Joanne B. Fertig. Gamma-glutamyltranspeptidase and carbohydrate deficient transferrin: alternative measures of excessive alcohol consumption. *Alcoholism, Clinical and Experimental Research*, 19(6):1541–6, dec 1995.

[158] Marsha Y. Morgan and Heather G. Major. The use of serum carbohydrate-deficient transferrin in the assessment of 'high risk offenders' in Great Britain. *Alcohol and Alcoholism*, 31(6):625–8, nov 1996.

[159] Rolf Iffland and F. Grassnack. [Epidemiologic study of CDT (carbohydrate-deficient transferrin) and other indicators of alcohol problems in high blood alcohol German automobile drivers]. *Blutalkohol*, 32(1):26–41, jan 1995.

[160] Anders Helander, Asgeir Husa, and Jan-Olof Jeppsson. Improved HPLC method for carbohydrate-deficient transferrin in serum. *Clinical Chemistry*, 49(11):1881–90, nov 2003.

[161] Jennifer P. Pascali, Federica Bortolotti, Daniela Sorio, Mariela Ivanova, Timothy M. Palmbach, and Franco Tagliaro. Improved capillary electrophoresis determination of carbohydrate-deficient transferrin including on-line immunosubtraction. *Medicine, Science, and the Law*, 51(1):26–31, jan 2011.

[162] Mauro Ferrari. *BioMEMS and Biomedical Nanotechnology: Volume II: Micro/Nano Technologies for Genomics and Proteomics*. Springer Science & Business Media, 2007.

[163] Jingli Yuan and Guilan Wang. Lanthanide complex-based fluorescence label for time-resolved fluorescence bioassay. *Journal of fluorescence*, 15(4):559–68, jul 2005.

[164] Gaye F. White, Konstantin L. Litvinenko, Stephen R. Meech, David L. Andrews, and Andrew J. Thomson. Multiphoton-excited luminescence of a lanthanide ion in a protein complex: Tb(3+) bound to transferrin. *Photochemical & Photobiological Sciences*, 3(1):47–55, jan 2004.

[165] Luca Bertini, Claudio Greco, Luca De Gioia, and Piercarlo Fantucci. Time-dependent density functional theory study of Fe2(CO)9 low-lying electronic excited states. *The journal of physical chemistry. A*, 110(47):12900–7, nov 2006.

[166] Nicholas Noinaj, Nicole C. Easley, Muse Oke, Naoko Mizuno, James Gumbart, Evzen Boura, Ashley N. Steere, Olga Zak, Philip Aisen, Emad Tajkhorshid, Robert W. Evans, Andrew R. Gorringe, Anne B. Mason, Alasdair C. Steven, and Susan K. Buchanan. Structural basis for iron piracy by pathogenic Neisseria. *Nature*, 483(7387):53–8, mar 2012.

[167] Guinevere Mathies, Peter Gast, Dennis N. Chasteen, Ashley N. Luck, Anne B. Mason, and Edgar J.J. Groenen. Exploring the Fe(III) binding sites of human serum transferrin with EPR at 275 GHz. *Journal of Biological Inorganic Chemistry*, 20(3):487–96, apr 2015.

[168] Frederick S. Richardson. Terbium(III) and europium(III) ions as luminescent probes and stains for biomolecular systems. *Chemical Reviews*, 82(5):541–552, oct 1982.

[169] Paul R. Selvin and John E. Hearst. Luminescence energy transfer using a terbium chelate: improvements on fluorescence energy transfer. *Proceedings of the National Academy of Sciences of the United States of America*, 91(21):10024–8, oct 1994.

[170] Sohrab Abdollahi and Wesley R. Harris. Determination of the Binding Constant of Terbium-Transferrin. *Iranian Journal of Chemistry and Chemical Engineering*, 25(1):45–52, apr 2006.

[171] Zhiguang Xiao and Anthony G. Wedd. The challenges of determining metal-protein affinities. *Natural product reports*, 27(5):768–89, may 2010.

[172] Alfarius Eko Nugroho and Hiroshi Morita. Circular dichroism calculation for natural products. *Journal of natural medicines*, 68(1):1–10, jan 2014.

[173] Wesley R Harris. Anion binding properties of the transferrins. Implications for function. *Biochimica et Biophysica Acta*, 1820(3):348–61, mar 2012.

[174] Fabian López-Vallejo, Thomas Caulfield, Karina Martínez-Mayorga, Marc A. Giulianotti, Adel Nefzi, Richard A. Houghten, and Jose L. Medina-Franco. Integrating virtual screening and combinatorial chemistry for accelerated drug discovery. *Combinatorial chemistry & high throughput screening*, 14(6):475–87, jul 2011.

[175] Yu-Chian Chen. Beware of docking! *Trends in pharmacological sciences*, 36(2):78–95, feb 2015.

[176] Elizabeth Yuriev, Mark Agostino, and Paul A. Ramsland. Challenges and advances in computational docking: 2009 in review. *Journal of Molecular Recognition*, 24(2):149–164, mar 2011.

[177] Dimitris K. Agrafiotis, Alan C. Gibbs, Fangqiang Zhu, Sergei Izrailev, and Eric Martin. Conformational sampling of bioactive molecules: a comparative study. *Journal of chemical information and modeling*, 47(3):1067–86.

[178] Sérgio F. Sousa, Pedro A. Fernandes, and Maria J. Ramos. Protein-ligand docking: current status and future challenges. *Proteins*, 65(1):15–26, oct 2006.

[179] Boris Gorelik and Amiram Goldblum. High quality binding modes in docking ligands to proteins. *Proteins*, 71(3):1373–86, may 2008.

[180] Nicolas Foloppe and Roderick Hubbard. Towards predictive ligand design with free-energy based computational methods? *Current medicinal chemistry*, 13(29):3583–608, jan 2006.

[181] Irwin D. Kuntz, Jeffrey M. Blaney, Stuart J. Oatley, Robert Langridge, and Thomas E. Ferrin. A geometric approach to macromolecule-ligand interactions. *Journal of molecular biology*, 161:269–288, 1982.

[182] Brian K. Shoichet, Irwin D. Kuntz, and Dale L. Bodian. Molecular docking using shape descriptors. *Journal of Computational Chemistry*, 13(3):380–397, apr 1992.

[183] 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.

[184] John A. Nelder and Roger Mead. A simplex-method for function minimization. *Computer Journal*, 7:308–313, 1964.

[185] Maria G. Cascio, Lisa A. Gauson, Lesley A. Stevenson, Ruth A. Ross, and Roger A. Pertwee. Evidence that the plant cannabinoid cannabigerol is a highly potent alpha2-adrenoceptor agonist and moderately potent 5HT1A receptor antagonist. *British journal of pharmacology*, 159(1):129–141, jan 2010.

[186] Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17):3389–402, sep 1997.

[187] David T. Jones. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, 292(2):195–202, sep 1999.

[188] Alexey Drozdetskiy, Christian Cole, James Procter, and Geoffrey J. Barton. JPred4: a protein secondary structure prediction server. *Nucleic acids research*, pages gkv332–, apr 2015.

[189] Jonathan J. Ward, Jaspreet S. Sodhi, Liam J. McGuffin, Brian F. Buxton, and David T. Jones. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of molecular biology*, 337(3):635–45, mar 2004.

[190] Narayanan Eswar, Ben Webb, Marc A. Marti-Renom, M. S. Madhusudhan, David Eramian, Min-Yi Shen, Ursula Pieper, and Andrej Sali. Comparative protein structure modeling using MODELLER. *Current protocols in protein science*, Chapter 2:Unit 2.9, nov 2007.

[191] Leigh Willard, Anuj Ranjan, Haiyan Zhang, Hassan Monzavi, Robert F. Boyko, Brian D. Sykes, and David S. Wishart. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic acids research*, 31(13):3316–3319, 2003.

[192] Paolo Mereghetti, Maria Luisa Ganadu, Elena Papaleo, Piercarlo Fantucci, and Luca De Gioia. Validation of protein models by a neural network approach. *BMC bioinformatics*, 9:66, jan 2008.

[193] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman J C Berendsen. GROMACS: fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–18, dec 2005.

[194] John L. Klepeis, Kresten Lindorff-Larsen, Ron O. Dror, and David E. Shaw. Long-timescale molecular dynamics simulations of protein structure and function. *Current opinion in structural biology*, 19(2):120–7, apr 2009.

[195] Berk Hess, Henk Bekker, Herman J.C. Berendsen, and Johannes Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18:1463–1472, 1997.

[196] Xavier Daura, Karl Gademann, Bernhard Jaun, Dieter Seebach, Wilfred F. van Gunsteren, and Alan E. Mark. Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie International Edition*, 38(1-2):236–240, jan 1999.

[197] Therese P. Lang, Scott R. Brozell, Sudipto Mukherjee, Eric F. Pettersen, Elaine C. Meng, Veena Thomas, Robert C. Rizzo, David A. Case, Thomas L. James, and Irwin D. Kuntz. DOCK 6: Combining techniques to model RNA-small molecule complexes. *RNA journal*, 15(6):1219–1230, 2009.

[198] Araz Jakalian, David B. Jack, and Christopher I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of computational chemistry*, 23(16):1623–41, dec 2002.

[199] Eric F. Pettersen, Thomas Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF Chimera: a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.

[200] Tom L. Blundell, B.L. Sibanda, Michael J.E. Sternberg, and Janet M. Thornton. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature*, 236:347–352, 1987.

[201] Jurgen Bajorath, Ronald Stenkamp, and Alejandro Aruffo. Knowledge-based model building of proteins: concepts and examples. *Protein Science*, 2:1798–1810, 1994.

[202] E.Y.T Chien, W. Liu, Q. Zhao, V. Katritch, G.W. Han, M.A. Hanson, L. Shi, A.H. Newman, J.A. Javitch, V. Cherezov, and R.C. Stevens. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science*, 330(6007):1091–1095, nov 2010.

[203] Cheng D. Wang, Melissa A. Buck, and Claire M. Fraser. Site-directed mutagenesis of a2A-adrenergic receptors: identification of amino acids involved in ligand binding and receptor activation by agonists. *Molecular Pharmacology*, 40:168–179, 1991.

[204] Jeffrey R. Jasper, John D. Lesnick, Katy L. Chang, Susan S. Yamanishi, Tae K. Chang, SherryA.O. Hsu, David A. Daunt, Douglas W. Bonhaus, and Richard M. Eglen. Ligand efficacy and potency at recombinant alpha 2 adrenergic receptors: agonist-mediated [35S]GTPgammaS binding. *Biochemical pharmacology*, 55:1035–1043, 1998.

[205] Tommi Nyronen, Marjo Pihlavisto, Juha M. Peltonen, Anna M Hoffren, Minna Varis, Tiina Salminen, Siegfrid Wurster, Anne Marjamaki, Liisa Kanerva, Erja Katainen, Leif Laaksonen, Juha M. Savola, Mika Scheinin, and Mark S. Johnson. Molecular Mechanism for Agonist-Promoted a2A -Adrenoceptor Activation by Norepinephrine and Epinephrine. *Molecular Pharmacology*, 59(5):1343–1354, 2001.

[206] Thomas A. Halgren, Robert B. Murphy, Richard A. Friesner, Hege S. Beard, Leah L. Frye, Thomas W. 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):1750–1759, 2004.

[207] William L. Jorgensen, David S. Maxwell, and Julian Tirado-Rives. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*, 7863(15):11225–11236, 1996.

[208] Richard A. Friesner, Robert B. Murphy, Matthew P. Repasky, Leah L. Frye, Jeremy R. Greenwood, Thomas A. Halgren, Paul C. Sanschagrin, and Daniel T. Mainz. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of medicinal chemistry*, 49(21):6177–96, oct 2006.

[209] Emmanuelle J. Meuillet. Novel inhibitors of AKT: assessment of a different approach targeting the pleckstrin homology domain. *Current Medicinal Chemistry*, 18(18):2727–42, jan 2011.

[210] Matthias P. Wymann and Carsten. Schultz. The chemical biology of phosphoinositide 3-kinases. *Chembiochem : a European journal of chemical biology*, 13(14):2022–35, sep 2012.

[211] Giuliana Cassinelli, Valentina Zuco, Laura Gatti, Cinzia Lanzi, Nadia Zaffaroni, Diego Colombo, and Paola Perego. Targeting the Akt kinase to modulate survival, invasiveness and drug resistance of cancer cells. *Current Medicinal Chemistry*, 20(15):1923–45, jan 2013.

[212] Francoise Chretien, Nicolas Moitessier, Fabien Roussel, J. P. Mauger, and Yves Chapleur. Carbohydrate-Based Mimics of D-myo-Inositol 1,4,5-Trisphosphate. *Current Organic Chemistry*, 4(5):513–534, may 2000.

[213] Laura Cipolla, Cristina Redaelli, Francesca Granucci, Giuseppe Zampella, Antonio Zaza, Riccardo Chisci, and Francesco Nicotra. Straightforward synthesis of novel Akt inhibitors based on a glucose scaffold. *Carbohydrate Research*, 345(10):1291–8, jul 2010.

[214] Luca Gabrielli, Ilaria Calloni, Giulia Donvito, Barbara Costa, Noemi Arrighetti, Paola Perego, Diego Colombo, Fiamma Ronchetti, Francesco Nicotra, and Laura Cipolla. Phosphatidylinositol 3-Phosphate Mimics Based on a Sulfoquinovose Scaffold: Synthesis and Evaluation as Protein Kinase B Inhibitors. *European Journal of Organic Chemistry*, 2014(27):5962–5967, sep 2014.

[215] Christine C. Milburn, Maria Deak, Sharon M. Kelly, Nick C. Price, Dario R. Alessi, and Daan M. F. van Aalten. Binding of phosphatidylinositol 3 , 4 , 5-trisphosphate to the pleckstrin homology domain of protein kinase B induces a conformational change. *Biochemical Journal*, 538:531–538, 2003.

[216] Véronique Calleja, Michel Laguerre, Peter J. Parker, and Banafshé Larijani. Role of a novel PH-kinase domain interface in PKB/Akt regulation: structural mechanism for allosteric inhibition. *PLoS biology*, 7(1):e17, jan 2009.

[217] Matteo Fumagalli, Manuela Sironi, Uberto Pozzoli, Anna Ferrer-Admettla, Linda Pattini, and Rasmus Nielsen. Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. *PLoS Genetics*, 7(11):e1002355, nov 2011.

[218] Manuela Sironi, Rachele Cagliani, Diego Forni, and Mario Clerici. Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nature reviews. Genetics*, 16(4):224–36, apr 2015.

[219] Carlos J. Camacho and Sandor Vajda. Protein-protein association kinetics and protein docking. *Current opinion in structural biology*, 12(1):36–40, feb 2002.

[220] Graham R. Smith and Michael J. E. Sternberg. Prediction of protein-protein interactions by docking methods. *Current opinion in structural biology*, 12(1):28–35, feb 2002.

[221] Sarel J. Fleishman, Jacob E. Corn, Eva M. Strauch, Tim A. Whitehead, Ingemar Andre, James Thompson, James J. Havranek, Rhiju Das, Philip Bradley, and David Baker. Rosetta in CAPRI rounds 13-19. *Proteins*, 78(15):3212–8, nov 2010.

[222] Dima Kozakov, David R. Hall, Dmitri Beglov, Ryan Brenke, Stephen R. Comeau, Yang Shen, Keyong Li, Jiefu Zheng, Pirooz Vakili, Ioannis C. Paschalidis, and Sandor Vajda. Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SDU, and stability analysis in CAPRI rounds 13-19. *Proteins*, 78(15):3124–30, nov 2010.

[223] Stephen R. Comeau, David W. Gatchell, Sandor Vajda, and Carlos J. Camacho. ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic acids research*, 32(Web Server issue):W96–9, jul 2004.

[224] Stephen R. Comeau, David W. Gatchell, Sandor Vajda, and Carlos J. Camacho. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*, 20(1):45–50, jan 2004.

[225] Dima Kozakov, Ryan. Brenke, Stephen R Comeau, and Sandor Vajda. PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins*, 65(2):392–406, nov 2006.

[226] Stephan Lorenzen and Yang Zhang. Identification of near-native structures by clustering protein docking conformations. *Proteins*, 68(1):187–94, jul 2007.

[227] Bernard R. Brooks, Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, and Martin Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2):187–217, 1983.

[228] Jeffrey J. Gray, Stewart Moughon, Chu Wang, Ora Schueler-Furman, Brian Kuhlman, Carol A. Rohl, and David Baker. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *Journal of molecular biology*, 331(1):281–99, aug 2003.

[229] Chu Wang, Ora Schueler-Furman, and David Baker. Improved side-chain modeling for protein-protein docking. *Protein Science*, 14(5):1328–39, may 2005.

[230] Tanja Kortemme, Alexandre V. Morozov, and David Baker. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *Journal of molecular biology*, 326(4):1239–59, feb 2003.

[231] Themis Lazaridis and Martin Karplus. Effective energy functions for protein structure prediction. *Current opinion in structural biology*, 10(2):139–45, apr 2000.

[232] Roland L. Dunbrack and Fred E. Cohen. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Science*, 6(8):1661–81, aug 1997.

[233] Lieping Chen. Co-inhibitory molecules of the B7-CD28 family in the control of T-cell immunity. *Nature reviews. Immunology*, 4(5):336–47, may 2004.

[234] Nargis Khan, Uthaman Gowthaman, Susanta Pahari, and Javed N. Agrewala. Manipulation of costimulatory molecules by intracellular pathogens: veni, vidi, vici!! *PLoS pathogens*, 8(6):e1002676, jan 2012.

[235] Mizuho Kajikawa, Pai-Chi Li, Eiji Goto, Naoyuki Miyashita, Masami Aoki-Kawasumi, Mari Mito-Yoshida, Mika Ikegaya, Yuji Sugita, and Satoshi Ishido. The intertransmembrane region of Kaposi's sarcoma-associated herpesvirus modulator of immune recognition 2 contributes to B7-2 downregulation. *Journal of virology*, 86(9):5288–96, may 2012.

[236] Dong Xu, Jian Zhang, Ambrish Roy, and Yang Zhang. Automated protein structure modeling in CASP9 by I-TASSER pipeline combined with QUARK-based ab initio folding and FG-MD-based structure refinement. *Proteins*, 79 Suppl 1:147–60, jan 2011.

[237] Sergey Lyskov and Jeffrey J. Gray. The RosettaDock server for local protein-protein docking. *Nucleic acids research*, 36(Web Server issue):W233–8, jul 2008.

[238] Sergey Lyskov, Fang-Chieh Chou, Shane O. Conchúir, Bryan S. Der, Kevin Drew, Daisuke Kuroda, Jianqing Xu, Brian D. Weitzner, Douglas P. Renfrew, Parin Sripakdeevong, Benjamin Borgo, James J. Havranek, Brian Kuhlman, Tanja Kortemme, Richard Bonneau, Jeffrey J. Gray, and Rhiju Das. Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE). *PloS one*, 8(5):e63906, jan 2013.

[239] Shaun M Lippow and Bruce Tidor. Progress in computational protein design. *Current Opinion in Biotechnology*, 18(4):305–11, aug 2007.

[240] Daniel M. Jordan, Vasily E. Ramensky, and Shamil R. Sunyaev. Human allelic variation: perspective from protein function, structure, and evolution. *Current Opinion in Structural Biology*, 20(3):342–50, jun 2010.

[241] Yana Bromberg and Burkhard Rost. Correlating protein function and stability through the analysis of single amino acid substitutions. *BMC bioinformatics*, 10 Suppl 8(8):S8, jan 2009.

[242] Yves Dehouck, Jean M. Kwasigroch, Dimitri Gilis, and Marianne Rooman. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC bioinformatics*, 12:151, jan 2011.

[243] Joost Schymkowitz, Jesper Borg, Francois Stricher, Robby Nys, Frederic Rousseau, and Luis Serrano. The FoldX web server: an online force field. *Nucleic acids research*, 33(Web Server issue):W382–8, jul 2005.

[244] Emidio Capriotti, Piero Fariselli, and Rita Casadio. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic acids research*, 33(Web Server issue):W306–10, jul 2005.

[245] Jacques Neefjes, Marlieke L M Jongsma, Petra Paul, and Oddmund Bakke. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature reviews. Immunology*, 11(12):823–36, dec 2011.

[246] Austin L. Hughes and Meredith Yeager. Natural selection at major histocompatibility complex loci of vertebrates. *Annual review of genetics*, 32:415–35, jan 1998.

[247] Ted H. Hansen and Marlene Bouvier. MHC class I antigen presentation: learning from viral evasion strategies. *Nature reviews. Immunology*, 9(7):503–13, jul 2009.

[248] Gracelin K. Tina, Rana Bhadra, and Narayanaswamy Srinivasan. PIC: Protein Interactions Calculator. *Nucleic acids research*, 35(Web Server issue):W473–6, jul 2007.

[249] Aida M. Andrés, Megan Y. Dennis, Warren W. Kretzschmar, Jennifer L. Cannons, Shih-Queen Lee-Lin, Belen Hurle, Pamela L. Schwartzberg, Scott H. Williamson, Carlos D. Bustamante, Rasmus Nielsen, Andrew G. Clark, and Eric D. Green. Balancing selection maintains a form of ERAP2 that undergoes nonsense-mediated decay and affects antigen presentation. *PLoS genetics*, 6(10):e1001157, oct 2010.

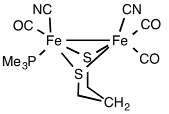
[250] Rachele Cagliani, Stefania Riva, Mara Biasin, Matteo Fumagalli, Uberto Pozzoli, Sergio Lo Caputo, Francesco Mazzotta, Luca Piacentini, Nereo Bresolin, Mario Clerici, and Manuela Sironi. Genetic diversity at endoplasmic reticulum aminopeptidases is maintained by balancing selection and is associated with natural resistance to HIV-1 infection. *Human molecular genetics*, 19(23):4705–14, dec 2010.

Appendix

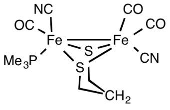
Chapter 2

Full set of models investigated in section 2.1.1.

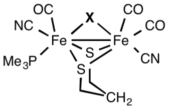
1



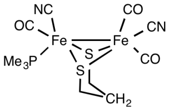
2



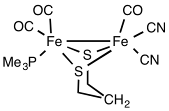
3



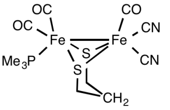
4



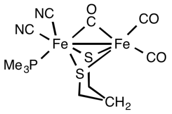
5



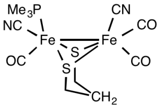
6



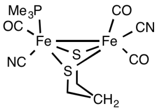
7



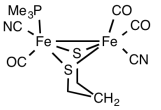
8



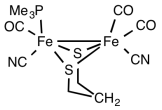
9



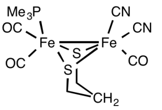
10



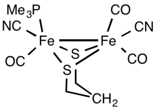
11



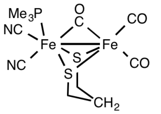
12



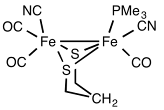
13



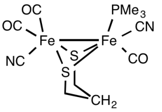
14



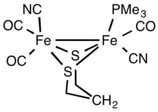
15



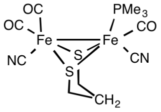
16



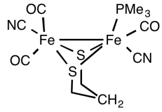
17



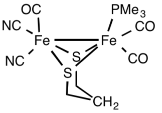
18



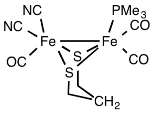
19



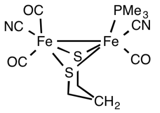
20



21



22



Chapter 4

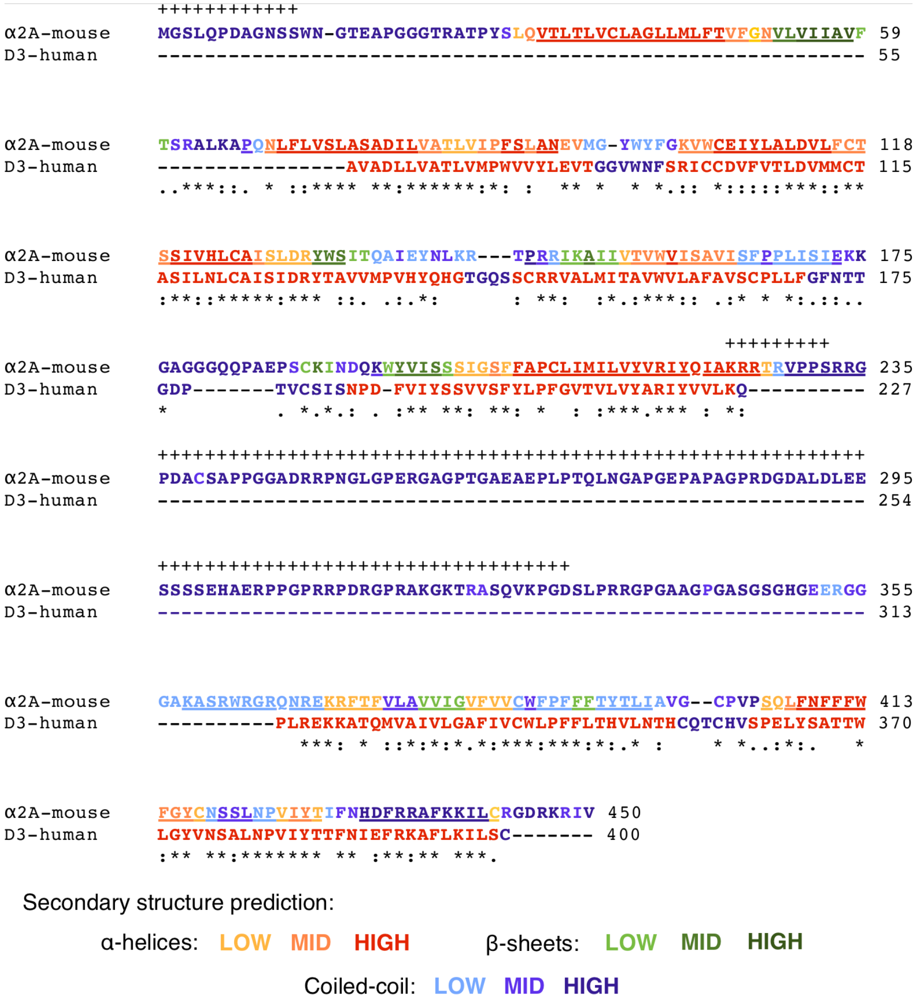
Multiple alignment of α_{2A}-ARs homologous sequences

```
sp|P35462|DRD3_HUMAN       ---------------MASLSQLSSHLNYTCGAENSTGASQA--RPHAY-----YALSYCA 38
sp|Q6TLI9|DRD2_MUSPF       ----------MDPLNLSWYDDDPESRNWSRPFNGSEGKVGK--PHYNY-----YAMLLTL 43
sp|Q64264|5HT1A_MOUSE      -------------MDMFSLGQGNNTTTSLEPFGTGGNDTGLSNVTFSYQVI--TSLLLGT 45
sp|P30545|ADA2B_MOUSE      ---------------------------------------MVHQEPYSVQATAAIASAITF 21
sp|Q01337|ADA2C_MOUSE      MASPALAAALAAAAAEGPNGSDAGEWGSGGGANASGTDWVPPPGQYSAGAVAGLAAVVGF 60
sp|Q90WY4|ADA2A_DANRE      --------------------------MICGANATNGT---NATKEYTLLVALPLSIAVGL 31
sp|Q01338|ADA2A_MOUSE      ---------------MGSLQPDAGNSSWNGTEAPGGG---TRATPYSLQVTLTLVCLAGL 42
sp|P08913|ADA2A_HUMAN      ---------------MGSLQPDAGNASWNGTEAPGGG---ARATPYSLQVTLTLVCLAGL 42
                                                                        .

sp|P35462|DRD3_HUMAN       LILAIVFGNGLVCMAVLKERALQTTTNYLVVSLAVADLLVATLVMPWVVYLEVTGGVWNF 98
sp|Q6TLI9|DRD2_MUSPF       LIFVIVFGNVLVCMAVSREKALQTTTNYLIVSLAVADLLVATLVMPWVVYLEVV-GEWKF 102
sp|Q64264|5HT1A_MOUSE      LIFCAVLGNACVVAAIALERSLQNVANYLIGSLAVTDLMVSVLVLPMAALYQVL-NKWTL 104
sp|P30545|ADA2B_MOUSE      LILFTIFGNALVILAVLTSRSLRAPQNLFLVSLAAADILVATLIIPFSLANELL-GYWYF 80
sp|Q01337|ADA2C_MOUSE      LIVFTVVGNVLVVIAVLTSRALRAPQNLFLVSLASADILVATLVMPFSLANELM-AYWYF 119
sp|Q90WY4|ADA2A_DANRE      LILLIIFGNVLVIIAVFTSRALRAPQNLFLVSLASADILVATLVMPFSLANELM-GMWTF 90
sp|Q01338|ADA2A_MOUSE      LMLFTVFGNVLVIIAVFTSRALKAPQNLFLVSLASADILVATLVIPFSLANEVM-GYWYF 101
sp|P08913|ADA2A_HUMAN      LMLLTVFGNVLVIIAVFTSRALKAPQNLFLVSLASADILVATLVIPFSLANEVM-GYWYF 101
                           *:.  :.**  *  *:  .::*:   * :: *** :*::*:.*::*     ::    * :

sp|P35462|DRD3_HUMAN       SRICCDVFVTLDVMMCTASILNLCAISIDRYTAVVMPVHYQHGTGQSSCRRVALMITAVW 158
sp|Q6TLI9|DRD2_MUSPF       SRIHCDIFVTLDVMMCTASILNLCAISIDRYTAVAMPMLYNTR--YSSKRRVTVMIAIVW 160
sp|Q64264|5HT1A_MOUSE      GQVTCDLFIALDVLCCTSSILHLCAIALDRYWAITDPIDYVNK---RTPRRAAALISLTW 161
sp|P30545|ADA2B_MOUSE      WRAWCEVYLALDVLFCTSSIVHLCAISLDRYWAVSRALEYNSK---RTPRRIKCIILTVW 137
sp|Q01337|ADA2C_MOUSE      GQVWCGVYLALDVLFCTSSIVHLCAISLDRYWSVTQAVEYNLK---RTPRRVKATIVAVW 176
sp|Q90WY4|ADA2A_DANRE      GGVWCEIYLALDVLFCTASITHLCAISLDRYWSITQAIEYNLK---RTPQRIKRIIFIVW 147
sp|Q01338|ADA2A_MOUSE      GKVWCEIYLALDVLFCTSSIVHLCAISLDRYWSITQAIEYNLK---RTPRRIKAIIVTVW 158
sp|P08913|ADA2A_HUMAN      GKAWCEIYLALDVLFCTSSIVHLCAISLDRYWSITQAIEYNLK---RTPRRIKAIIITVW 158
                               * ::::***: **:** .****::*** ::   : *       : :*    *  .*

sp|P35462|DRD3_HUMAN       VLAFAVSCPLLFGFNTTGD-------PTVCSISNP-DFVIYSSVVSFYLPFGVTVLVYAR 210
sp|Q6TLI9|DRD2_MUSPF       VLSFTISCPLLFGLNNT-D-------QNECIIANP-AFVVYSSVVSFYVPFIVTLLVYIK 211
sp|Q64264|5HT1A_MOUSE      LIGFLISIPPMLGWRTPEDRS----NPNECTISKDHGYTIYSTFGAFYIPLLLMLVLYGR 217
sp|P30545|ADA2B_MOUSE      LIAAVISLPPLIYKG---DQRPEPHGLPQCELNQEAWYILASSIGSFFAPCLIMILVYLR 194
sp|Q01337|ADA2C_MOUSE      LISAVISFPPLVSFYRRPDGA----AYPQCGLNDETWYILSSCIGSFFAPCLIMGLVYAR 232
sp|Q90WY4|ADA2A_DANRE      IIAAVISCPPLITMKKSE--------GDICDINKEKWYIVSSCIGSFFLPCIIMVLVYIR 199
sp|Q01338|ADA2A_MOUSE      VISAVISFPPLISIEKKGAGGGQQPAEPSCKINDQKWYVISSSIGSFFAPCLIMILVYVR 218
sp|P08913|ADA2A_HUMAN      VISAVISFPPLISIEKKGGGGGPQPAEPRCEINDQKWYVISSCIGSFFAPCLIMILVYVR 218
                           ::.  :* * :.                 * : .   : : * . :*: *  :  ::* :

sp|P35462|DRD3_HUMAN       IYVVLKQRRRKRILTRQNSQCNSVRPGFPQQTLSPDPA------HLEL------------ 252
sp|Q6TLI9|DRD2_MUSPF       IYIVLRRRRKRVNTKR-SSRAFRANLKAPLKGNYTHPE------DMKLCTV--IMKSNGS 262
sp|Q64264|5HT1A_MOUSE      IFRAARFRIRKTVKKVEKKGAGTSFGTSSAP---------------------PPKKSLNG 256
sp|P30545|ADA2B_MOUSE      IYVIAKRSHCRGLGAKRGSG------EGESKKPRPGPAAGGVPASAKVPTLVSPLSSVGE 248
sp|Q01337|ADA2C_MOUSE      IYRVAKLRTRTLSEKRGPAGP---DGASPTTENGLGKAAGEN--G--------------- 272
sp|Q90WY4|ADA2A_DANRE      IYQIAKKRTRAPPGDHRKNEVGKKENDPHEK--------------------------LNG 233
sp|Q01338|ADA2A_MOUSE      IYQIAKRRTRVPPSRRGPDACSAPPGGADRRPNGLGPERGAGPTGAEAEPLPTQ---LNG 275
sp|P08913|ADA2A_HUMAN      IYQIAKRRTRVPPSRRGPDAVAAPPGGTERRPNGLGPERSAGPGGAEAEPLPTQ---LNG 275
                           *:   :
```

```
sp|P35462|DRD3_HUMAN       ------------------------------KRYYSICQ---DTALGGP---GFQERG--- 273
sp|Q6TLI9|DRD2_MUSPF       FPVNRRRVEAARRAQELEMEMLSSTSPPERTRYSPIPPSHHQLTLPDPSHHGLHSTA--- 319
sp|Q64264|5HT1A_MOUSE      QPG---SGDCRRSA-------------E-------------NRAVGTPCANGAVRQG--- 284
sp|P30545|ADA2B_MOUSE      ANG---HPKPPREKEEGETPEDP------------------EARALPPNWSALPRSVQDQ 287
sp|Q01337|ADA2C_MOUSE      ------HCAPPR----TEVEPDESSA-A-------------ERRRRRGA----LRRGGRR 304
sp|Q90WY4|ADA2A_DANRE      IQ----NAEPDDKDEINGVDMEESSS-S-------------DHKVSNPC----SLK---K 268
sp|Q01338|ADA2A_MOUSE      APG---EPAPAGPRDGDALDLEESSS-S-------------EHAERPPG----PRRPDRG 314
sp|P08913|ADA2A_HUMAN      APG---EPAPAGPRDTDALDLEESSS-S-------------DHAERPPG----PRRPERG 314
                                                                    :

sp|P35462|DRD3_HUMAN       ----------GELK--------------REEKTRNSLSPTIAPKLSLE--------VRKL 301
sp|Q6TLI9|DRD2_MUSPF       ---------NSPVK--------------PEKNGHAKDHPKIA--KIFE--------IQSM 346
sp|Q64264|5HT1A_MOUSE      --------------EDDATLEVIEVHRVGNSKGHL-----PLPSESGATSY-----VPA- 319
sp|P30545|ADA2B_MOUSE      KKGTSGATAEKGAEEDEEEVEECEPQTLPASPASVFNPPLQQPQTSRVLATL--RG-QVL 344
sp|Q01337|ADA2C_MOUSE      REGAEGD-TGSADG---------PGPGLAAEQGART--ASRSPGPGGRLSRASSRSVEFF 352
sp|Q90WY4|ADA2A_DANRE      K-SSKGKTKLSQIK---------PGDG---------------------------D----- 286
sp|Q01338|ADA2A_MOUSE      P-RAKGKTRASQVK---------PGDSLPR----------RGPGAAGPGASGSGH----- 349
sp|P08913|ADA2A_HUMAN      P-RGKGKARASQVK---------PGDSLPR----------RGPGATGIGTPAAGP----- 349


sp|P35462|DRD3_HUMAN       SNGRLSTSLKLGPLQPRGVPLREKKATQMVAIVLGAFIVCWLPFFLTHVLNTHC-QTCHV 360
sp|Q6TLI9|DRD2_MUSPF       PNGKTRTSLKTMSR-RKLSQQKEKKATQMLAIVLGVFIICWLPFFITHILNIHC-D-CNI 403
sp|Q64264|5HT1A_MOUSE      --CLERKNERTAEAKRKMALARERKTVKTLGIIMGTFILCWLPFFIVALVLPFCESSCHM 377
sp|P30545|ADA2B_MOUSE      L-SKNVGVASGQWWRRRTQLSREKRFTFVLAVVIGVFVVCWFPFFFSYSLGAICPQHCKV 403
sp|Q01337|ADA2C_MOUSE      L-SRR-RRARSSVCRRKVAQAREKRFTFVLAVVMGVFVLCWFPFFFSYSLYGICREACQL 410
sp|Q90WY4|ADA2A_DANRE      --KTE-ACQTTKASRWKGRQNREKRFTFVLAVVIGVFVICWFPFFFTYTFTAFC--DCCV 341
sp|Q01338|ADA2A_MOUSE      --GEE-RGGGAKASRWRGRQNREKRFTFVLAVVIGVFVVCWFPFFFTYTLIA-V--GCPV 403
sp|P08913|ADA2A_HUMAN      --GEE-RVGAAKASRWRGRQNREKRFTFVLAVVIGVFVVCWFPFFFTYTLTA-V--GCSV 403
                                           :    :*:: .  :.:::*.*::**:***:   .       * :

sp|P35462|DRD3_HUMAN       SPELYSATTWLGYVNSALNPVIYTTFNIEFRKAFLKILSC--------     400
sp|Q6TLI9|DRD2_MUSPF       PPVLYSAFTWLGYVNSAVNPIIYTTFNVEFRKAFMKILHC--------     443
sp|Q64264|5HT1A_MOUSE      PELLGAIINWLGYSNSLLNPVIYAYFNKDFQNAFKKIIKCKFCR----     421
sp|P30545|ADA2B_MOUSE      PHGLFQFFFWIGYCNSSLNPVIYTIFNQDFRRAFRRILCRQWTQTGW-     450
sp|Q01337|ADA2C_MOUSE      PEPLFKFFFWIGYCNSSLNPVIYTVFNQDFRRSFKHILFRRRRRGFRQ     458
sp|Q90WY4|ADA2A_DANRE      PETLFKFFFWFGYCNSSLNPIIYTIFNNDFRRSFKKILCRRDKRRVV-     388
sp|Q01338|ADA2A_MOUSE      PSQLFNFFFWFGYCNSSLNPVIYTIFNHDFRRAFKKILCRGDRKRIV-     450
sp|P08913|ADA2A_HUMAN      PRTLFKFFFWFGYCNSSLNPVIYTIFNHDFRRAFKKILCRGDRKRIV-     450
                              *     *:** ** :**:**: ** :*:.:* :*:
```

Manually-refined sequence alignment between murine α_{2A} isoform and human D3 receptor.

For α_{2A}-AR the probability of a certain secondary structure predicted by PSIPRED is represented by a colour scale (from yellow to red for a-helices, from light green to dark green for β-sheets and from light blue to blue for coiled coil); α-helices predicted by JPRED are indicated as underlining of α_{2A}-AR sequence and disordered regions predicted by DISOPRED are highlighted through cross-shaped symbols over the residues. '*' residues are identical in all sequences of the allignment, ':' conserved substitutions, '.' semi-conserved substitutions.



```
             ++++++++++++
α2A-mouse    MGSLQPDAGNSSWN-GTEAPGGGTRATPYSLQVTLTLVCLAGLLMLFTVFGNVLVIIAVF 59
D3-human     ------------------------------------------------------------ 55


α2A-mouse    TSRALKAPQNLFLVSLASADILVATLVIPFSLANEVMG-YWYFGKVWCEIYLALDVLFCT 118
D3-human     ----------------AVADLLVATLVMPWVVYLEVTGGVWNFSRICCDVFVTLDVMMCT 115
             ..***::. * ::**** **:******:*: :  ** *  * *.:: *:::::***::**

α2A-mouse    SSIVHLCAISLDRYWSITQAIEYNLKR---TPRRIKAIIVTVWVISAVISFPPLISIEKK 175
D3-human     ASILNLCAISIDRYTAVVMPVHYQHGTGQSSCRRVALMITAVWVLAFAVSCPLLFGFNTT 175
             :**::*****:*** ::. .:.*:      : **:  :*.:***:: .:* * *:.::..

                                                            +++++++++
α2A-mouse    GAGGGQQPAEPSCKINDQKWYVISSSIGSFFAPCLIMILVYVRIYQIAKRRTRVPPSRRG 235
D3-human     GDP-------TVCSISNPD-FVIYSSVVSFYLPFGVTVLVYARIYVVLKQ---------- 227
             *         . *.*.: . :** **: **: *  : :***.*** : *:

             ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
α2A-mouse    PDACSAPPGGADRRPNGLGPERGAGPTGAEAEPLPTQLNGAPGEPAPAGPRDGDALDLEE 295
D3-human     ------------------------------------------------------------ 254


             +++++++++++++++++++++++++++++++++++
α2A-mouse    SSSSEHAERPPGPRRPDRGPRAKGKTRASQVKPGDSLPRRGPGAAGPGASGSGHGEERGG 355
D3-human     ------------------------------------------------------------ 313


α2A-mouse    GAKASRWRGRQNREKRFTFVLAVVIGVFVVCWFPFFFTYTLIAVG--CPVPSQLFNFFFW 413
D3-human     ----------PLREKKATQMVAIVLGAFIVCWLPFFLTHVLNTHCQTCHVSPELYSATTW 370
                         ***: * ::*:*:*.*:***:***:*:.* :    * *..:*:.   *

α2A-mouse    FGYCNSSLNPVIYTIFNHDFRRAFKKILCRGDRKRIV 450
D3-human     LGYVNSALNPVIYTTFNIEFRKAFLKILSC------- 400
             :** **:******* ** :**:** ***.
```

Model quality assessment: Vadar

Model 1

Statistic	Observed	Expected
Residues in phipsi core	424 (94%)	405 (90%)
Residues in phipsi allowed	23 (5%)	32 (7%)
Residues in phipsi generous	1 (0%)	5 (1%)
Residues in phipsi outside	0 (0%)	0 (0%)
Residues in omega core	442 (98%)	432 (96%)
Residues in omega allowed	7 (1%)	13 (3%)
Residues in omega generous	1 (0%)	0 (0%)
Residues in omega outside	0 (0%)	5 (1%)
Packaging defects	239	31
Free energy of folding	-324.26	-429.48
Residues 95% buried	64	181
Buried charges	7	0

Model 2

Statistic	Observed	Expected
Residues in phipsi core	422 (93%)	405 (90%)
Residues in phipsi allowed	22 (4%)	32 (7%)
Residues in phipsi generous	4 (0%)	5 (1%)
Residues in phipsi outside	0 (0%)	0 (0%)
Residues in omega core	447 (99%)	432 (96%)
Residues in omega allowed	1 (0%)	13 (3%)
Residues in omega generous	1 (0%)	0 (0%)
Residues in omega outside	0 (0%)	5 (1%)
Packaging defects	187	31
Free energy of folding	-314.74	-429.48
Residues 95% buried	64	181
Buried charges	7	0

Model 3

Statistic	Observed	Expected
Residues in phipsi core	422 (93%)	405 (90%)
Residues in phipsi allowed	24 (5%)	32 (7%)
Residues in phipsi generous	3 (0%)	5 (1%)
Residues in phipsi outside	0 (0%)	0 (0%)
Residues in omega core	444 (98%)	432 (96%)
Residues in omega allowed	5 (1%)	13 (3%)
Residues in omega generous	1 (0%)	0 (0%)
Residues in omega outside	0 (0%)	5 (1%)
Packaging defects	223	31
Free energy of folding	-332.41	-429.48
Residues 95% buried	68	181
Buried charges	4	0

Model 4

Statistic	Observed	Expected
Residues in phipsi core	421 (93%)	405 (90%)
Residues in phipsi allowed	20 (4%)	32 (7%)
Residues in phipsi generous	7 (1%)	5 (1%)
Residues in phipsi outside	0 (0%)	0 (0%)
Residues in omega core	429 (95%)	432 (96%)
Residues in omega allowed	14 (3%)	13 (3%)
Residues in omega generous	4 (0%)	0 (0%)
Residues in omega outside	2 (0%)	5 (1%)
Packaging defects	84	31
Free energy of folding	-325.48	-429.48
Residues 95% buried	71	181
Buried charges	5	0

Model 5

Statistic	Observed	Expected
Residues in phipsi core	420 (93%)	405 (90%)
Residues in phipsi allowed	24 (5%)	32 (7%)
Residues in phipsi generous	3 (0%)	5 (1%)
Residues in phipsi outside	1 (0%)	0 (0%)
Residues in omega core	440 (97%)	432 (96%)
Residues in omega allowed	8 (1%)	13 (3%)
Residues in omega generous	2 (0%)	0 (0%)
Residues in omega outside	0 (0%)	5 (1%)
Packaging defects	249	31
Free energy of folding	-314.45	-429.48
Residues 95% buried	69	181
Buried charges	7	0

Model 6

Statistic	Observed	Expected
Residues in phipsi core	422 (93%)	405 (90%)
Residues in phipsi allowed	22 (4%)	32 (7%)
Residues in phipsi generous	3 (0%)	5 (1%)
Residues in phipsi outside	1 (0%)	0 (0%)
Residues in omega core	445 (98%)	432 (96%)
Residues in omega allowed	4 (0%)	13 (3%)
Residues in omega generous	1 (0%)	0 (0%)
Residues in omega outside	0 (0%)	5 (1%)
Packaging defects	78	31
Free energy of folding	-311.28	-429.48
Residues 95% buried	63	181
Buried charges	5	0

Model 7

Statistic	Observed	Expected
Residues in phipsi core	426 (94%)	405 (90%)
Residues in phipsi allowed	18 (4%)	32 (7%)
Residues in phipsi generous	4 (0%)	5 (1%)
Residues in phipsi outside	1 (0%)	0 (0%)
Residues in omega core	442 (98%)	432 (96%)
Residues in omega allowed	5 (1%)	13 (3%)
Residues in omega generous	2 (0%)	0 (0%)
Residues in omega outside	0 (0%)	5 (1%)
Packaging defects	226	31
Free energy of folding	-314.60	-429.48
Residues 95% buried	62	181
Buried charges	6	0

Model 8

Statistic	Observed	Expected
Residues in phipsi core	422 (93%)	405 (90%)
Residues in phipsi allowed	23 (5%)	32 (7%)
Residues in phipsi generous	5 (1%)	5 (1%)
Residues in phipsi outside	0 (0%)	0 (0%)
Residues in omega core	441 (98%)	432 (96%)
Residues in omega allowed	8 (1%)	13 (3%)
Residues in omega generous	1 (0%)	0 (0%)
Residues in omega outside	0 (0%)	5 (1%)
Packaging defects	88	31
Free energy of folding	-331.37	-429.48
Residues 95% buried	63	181
Buried charges	10	0

Model 9

Statistic	Observed	Expected
Residues in phipsi core	425 (94%)	405 (90%)
Residues in phipsi allowed	21 (4%)	32 (7%)
Residues in phipsi generous	3 (0%)	5 (1%)
Residues in phipsi outside	0 (0%)	0 (0%)
Residues in omega core	443 (98%)	432 (96%)
Residues in omega allowed	6 (1%)	13 (3%)
Residues in omega generous	1 (0%)	0 (0%)
Residues in omega outside	0 (0%)	5 (1%)
Packaging defects	196	31
Free energy of folding	-316.42	-429.48
Residues 95% buried	63	181
Buried charges	4	0

Model 10

Statistic	Observed	Expected
Residues in phipsi core	419 (93%)	405 (90%)
Residues in phipsi allowed	25 (5%)	32 (7%)
Residues in phipsi generous	4 (0%)	5 (1%)
Residues in phipsi outside	0 (0%)	0 (0%)
Residues in omega core	442 (98%)	432 (96%)
Residues in omega allowed	8 (1%)	13 (3%)
Residues in omega generous	0 (0%)	0 (0%)
Residues in omega outside	0 (0%)	5 (1%)
Packaging defects	87	31
Free energy of folding	-317.35	-429.48
Residues 95% buried	65	181
Buried charges	4	0

Model quality assessment: Molpdf and DOPE score

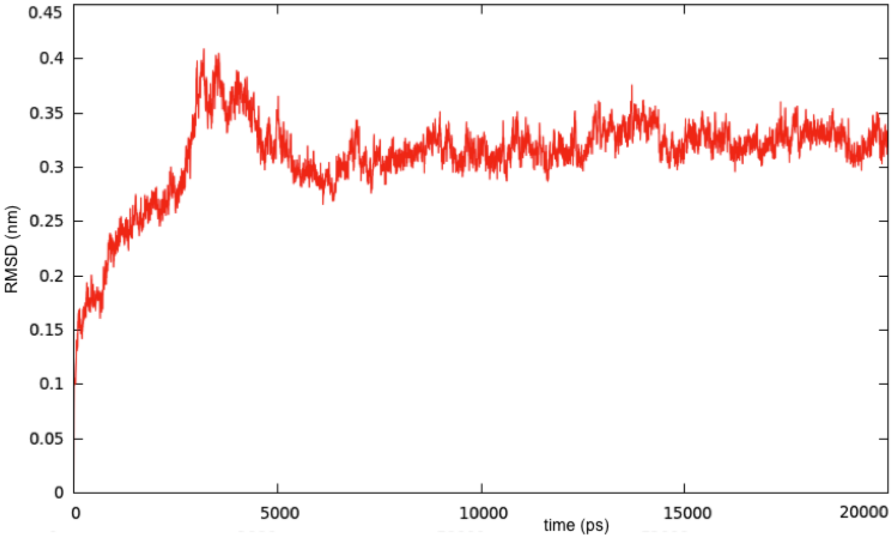
	molpdf	DOPE score
Model 1	2145.63452	-42204.50391
Model 2	2208.24219	-42659.00781
Model 3	2242.48047	-42120.86328
Model 4	2374.85205	-42462.08203
Model 5	2506.38184	-42400.53125
Model 6	2731.14307	-42650.35547
Model 7	2166.48169	-42428.48828
Model 8	2223.27759	-42544.02344
Model 9	2227.86157	-42537.24609
Model 10	2373.13159	-42302.64453

Model quality assessment: Aide

	Predicted LG-score	Predicted TM-score
Model 1	0.2170685	0.61
Model 2	0.2210268	0.63
Model 3	0.2178228	0.64
Model 4	0.1943737	0.64
Model 5	0.2037580	0.63
Model 6	0.2004316	0.62
Model 7	0.2121720	0.64
Model 8	0.2147148	0.64
Model 9	0.2018273	0.64
Model 10	0.1985326	0.64

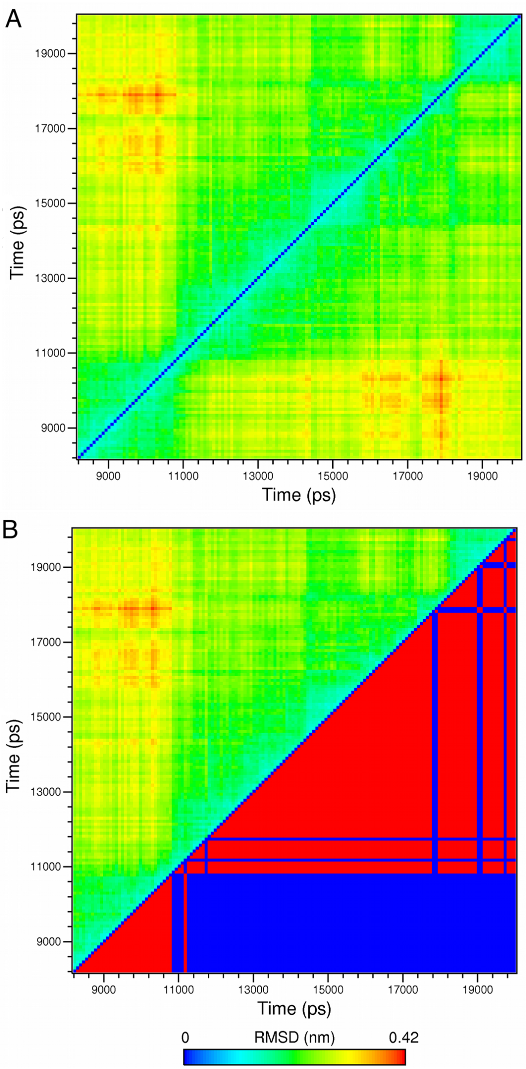
RMSD

RMSD profile of the murine α_{2A}-ARs trajectory.



RMSD matrices and clustering

RMSD matrix of the trajectory (panel A) and clustering analysis performed on the RMSD matrix (panel B).



List of publications

The thesis is based on the following publications (attached below):

- Silvia Nicotra, **Giulia Filippi**, Rita Grandori, Luca De Gioia, Carlo Santambrogio, Franco Tagliaro. Transferrin-terbium fluorescent complex for the diagnosis of alcohol abuse. In preparation.
- **Giulia Filippi**, Francesca Comelli, Elena Papaleo, Roger Pertwee, Luca De Gioia, and Barbara Costa. The role of alpha2-adrenoceptors in cannabigerol-induced antinociception: a computational and a pharmacological study. In preparation.
- Barbara Costa, Milind Dangate, Maria Vetro, Giulia Donvito, Luca Gabrielli, Loredana Amigoni, Giuliana Cassinelli, Cinzia Lanzi, Michela Ceriani, **Giulia Filippi**, Laura Cipolla, Nadia Zaffaroni, Paola Perego, Diego Colombo. Synthetic sulfoglycolipids targeting the serine-threonine protein kinase Akt. *ChemMedChem*, Submitted.
- **Giulia Filippi**, Federica Arrigoni, Luca Bertini, Luca De Gioia, and Giuseppe Zampella. DFT dissection of the reduction step in H2 catalytic production by [FeFe]-hydrogenase-inspired models: can the bridging hydride become more reactive than the terminal isomer? *Inorganic Chemistry*, 54(19):9529-9542, 2015.
- Diego Forni, Rachele Cagliani, Claudia Tresoldi, Uberto Pozzoli, Luca De Gioia, **Giulia Filippi**, Stefania Riva, Giorgia Menozzi, Marta Colleoni, Mara Biasin, Sergio Lo Caputo, Francesco Mazzotta, Giacomo P. Comi, Nereo Bresolin, Mario Clerici, and Manuela Sironi. An evolutionary analysis of antigen processing and presentation across different timescales reveals pervasive selection. *PLoS genetics*, 10(3):e1004189, 2014.
- Sabrina Sicolo, Maurizio Bruschi, Luca Bertini, Giuseppe Zampella, **Giulia Filippi**, Federica Arrigoni, Luca De Gioia, and Claudio Greco. Towards biomimetic models of the reduced [FeFe]-hydrogenase that preserve the key structural features of the enzyme active site; a DFT investigation. *International Journal of Hydrogen Energy*, 39(32):18565–18573, 2014.
- Diego Forni, Rachele Cagliani, Uberto Pozzoli, Marta Colleoni, Stefania Riva, Mara Biasin, **Giulia Filippi**, Luca De Gioia, Federica Gnudi, Giacomo P. Comi, Nereo

Bresolin, Mario Clerici, and Manuela Sironi. A 175 Million Year History of T Cell Regulatory Molecules Reveals Widespread Selection, with Adaptive Evolution of Disease Alleles. *Immunity*, 38:1129-41.

Additional publications not discussed in this thesis:

- Rachele Cagliani, Diego Forni, **Giulia Filippi**, Alessandra Mozzi, Luca De Gioia, Chiara Pontremoli, Uberto Pozzoli, Nereo Bresolin, Mario Clerici, and Manuela Sironi. The complement system as an epitome of host-pathogen genetic conflicts. *Molecular Ecology*, Accepted.
- Diego Forni, **Giulia Filippi**, Rachele Cagliani, Luca De Gioia, Uberto Pozzoli, Nasser Al-Daghri, Mario Clerici, and Manuela Sironi. The heptad repeat region is a major selection target in MERS-CoV and related coronaviruses. *Scientific Reports*, 5:14480, 2015.
- Marco Rigamonti, Silvia Groppi, Fiorella Belotti, Roberto Ambrosini, **Giulia Filippi**, Enzo Martegani, and Renata Tisi. Hypotonic stress-induced calcium signaling in Saccharomyces cerevisiae involves TRP-like transporters on the endoplasmic reticulum membrane. *Cell Calcium*, 57(2):57–68, 2015.
- Rachele Cagliani, Diego Forni, Claudia Tresoldi, Uberto Pozzoli, **Giulia Filippi**, Veronica Rainone, Luca De Gioia, Mario Clerici, and Manuela Sironi. RIG-I-like receptors evolved adaptively in mammals, with parallel evolution at LGP2 and RIG-I. *Journal of Molecular Biology*, 426(6):1351–65, 2014.
- Manuela Sironi, Mara Biasin, Rachele Cagliani, Federica Gnudi, Irma Saulle, Salomè Ibba, **Giulia Filippi**, Sarah Yahyaei, Claudia Tresoldi, Stefania Riva, Daria Trabattoni, Luca De Gioia, Sergio Lo Caputo, Francesco Mazzotta, Diego Forni, Chiara Pontremoli, Juan Antonio Pineda, Uberto Pozzoli, Antonio Rivero-Juarez, Antonio Caruz and Mario Clerici. Evolutionary Analysis Identifies an MX2 Haplotype Associated with Natural Resistance to HIV-1 Infection. *Molecular Biology and Evolution*, 31(9):2402–14, 2014.


carbohydrate deficient transferrin
ROSIE
hydrogenases
MODELLER
medicinal chemistry
scoring function
CN-
E°
structural
COMPUTATIONAL
Clostridium pasteurianum
wavefunction
adrenoceptors
dynamic
pKa
REDOX
UV-Vis
ΔG°
equation
hydrogen bond
Kd
PoPMuSiC
LUMO
HCOOH
terbium
catalysis
evolution
spectra
PMe3
F=ma
cannabigerol
interactions
ClusPro
NHC
aminoacid
models
CDT
algorithm
DFT
Schrödinger
docking
CO2
RMSD
geometry
optimiztion
ΔE
proteins
alcohol abuse
fluorescence
minimization
solvent
hydrogen