Università di Milano-Bicocca
Facoltà di Scienze Matematiche, Fisiche e Naturali
Dipartimento di Biotecnologie e Bioscienze
XXVI Ciclo di Dottorato in Scienze Chimiche

Computational investigation of structure-function relationship in metalloenzymes

Coordinatore: Prof. Giorgio Moro
Tutor: Dott. Giuseppe Zampella
Cotutor: Dott.ssa Elena Papaleo

Matteo Tiberti, 057957

Abstract

Important processes such as carbon dioxide capture, hydrogen evolution and oxidation as well as bioremediation are of paramount importance for the development of an environmental-friendly economy. In nature, the reactions underlying these processes are naturally performed by metalloenzymes. The understanding of how they work will ease the development of important technological and biotechnological applications, such as bio-inspired inorganic catalysts or more effective enzymes. In the present work, we have used computational methods to study structure-function relationships in metalloenzymes which catalyze important reactions involved in environment-friendly processes. In particular, we selected three cases of study; the formate dehydrogenase (FDH) from *Escherichia coli*, which catalyzes the conversion between carbon dioxide and formate, the [NiFe]-hydrogenase from *Allochromatium vinosum*, which performs hydrogen oxidation, and the phosphotriesterase (PTE) from *Agrobacterium radiobacter* that is able to hydrolyse organophosphates. Since the study of structure-function relationship in enzymes generally requires both a view on the reaction mechanism and on the functional dynamics, we had to employ different levels of theory. At first, we focused on the study of metallo-enzymes for which the catalytic mechanisms still have to be clarified in details, as FDH and a [NiFe]-hydrogenase. In particular, we studied the oxidation of formate to carbon dioxide at the Mo-binding cofactor in the active site of FDH, using Density Functional Theory (DFT). We employed for the first time a cluster model for the molybdopterin cofactor which is meant to best reproduce the stereoelectronic properties of the whole moiety. By testing several reaction pathways for the rate-determining step of FDH catalytic cycle, we were able to propose a novel

reaction mechanism which includes a β-hydride elimination step with a metal hydride intermediate. We then employed similar methods to study hydrogen oxidation in [NiFe]-hydrogenase. We have investigated the dihydrogen coordination mode and subsequent oxidation, using model clusters of different sizes. It turned out that the spin state and the distorted see-saw coordination geometry of the Ni ion are two crucial factors that tune the energetics and regiochemistry of H_2 binding. We were able to propose a reaction pathway involving the oxidative addition of H_2 followed by proton transfer to the sulfur atom of one of the terminally coordinated cysteines as the lowest energy one. We then focused on another metalloenzyme, in which experimental evidences support a role of conformational dynamics for the catalysis, PTE. In this enzyme, it has been shown that the substrate hydrolysis can occur on fast time-scales, whereas one of the rate limiting steps seems to be the transition between so-called "open" and "closed" conformations with respect to the metal-binding site. Indeed, in PTE mutations of residues, which are not directly involved in the active site, are known to affect the population of the closed and open states, as well as to alter the k_{cat}. We thus studied this enzyme by atomistic Molecular Dynamics (MD) simulations to investigate the long-range structural communication routes between mutation sites and active site residues. We did so by considering both wild-type PTE (arPTE-WT) and a multiple mutant (arPTE-8M) in which mutations alter the equilibrium between the “closed” and “open” populations. By employing both freely available and in-house produced MD trajectories analysis tools, we were able to identify the major communication routes in the enzymes as well as to point out which mutation sites are more likely to be involved in the transmission of structural information. Comparing the communication pathways in the two variants allowed us to investigate how the presence of mutations influence the communication pathways in the protein and to relate them to the effects induce by the mutations on protein function. In conclusion, we have studied function-structure relationships in three different metalloenzymes, FDH, [NiFe] hydrogenase and PTE, by using high level DFT or MD methods and developing ad-hoc analysis tools when required.

Contents

Publications

The present thesis has led to the production of the following scientific papers:

Evidence for the formation of a Mo-H intermediate in the catalytic cycle of formate dehydrogenase. Matteo Tiberti, Elena Papaleo, Nino Russo, Luca De Gioia, Giuseppe Zampella§. Inorganic Chemistry, 51 (15), 8331-8339, 2012

Disclosure of key stereoelectronic factors for efficient H_2 binding and cleavage in the active site of [NiFe]-hydrogenases. Maurizio Bruschi, Matteo Tiberti, Alessandro Guerra, Luca De Gioia§, Journal of the American Chemical Society, 136 (5), 1803–1814, 2014

Dynamics properties of a bacterial phosphotriesterase and its mutants investigated by molecular dynamics simulations and protein structure networks. Matteo Tiberti, Kresten Lindorff-Larsen, Nino Russo, Giuseppe Zampella, Luca De Gioia, Elena Papaleo§, *To be submitted*

Other works realized during the PhD period:

Communication Routes in ARID Domains between Distal Residues in Helix 5 and the DNA-Binding Loops. Gaetano Invernizzi, Matteo Tiberti, Matteo Lambrughi, Kresten Lindorff-Larsen, Elena Papaleo§, PLoS Computational Biology, 2014; 10 (9), e1003744.

PyInteraph: a framework for the analysis of interaction networks in structural ensembles of proteins. Matteo Tiberti§, Gaetano Invernizzi, Matteo Lambrughi, Yuval Inbar, Gideon Schreiber, Elena Papaleo§, Journal of Chemical Information and Modeling, (54) 5, 1537–1551, 2014

Functional annotation of the mesophilic-like character of mutants in a cold-adapted enzyme by self-organising map analysis of their molecular dynamics. Domenico Fraccalvieri, Matteo Tiberti, Alessandro Pandini,

Laura Bonati[§] , Elena Papaleo[§], Molecular BioSystems, 8 (10), 2680-2691, 2012

xPyder: A PyMOL Plugin To Analyze Coupled Residues and Their Networks in Protein Structures. Marco Pasi*, Matteo Tiberti*, Alberto Arrigoni, Elena Papaleo[§], Journal of Chemical Information and Modeling, 52 (7), 1865-1874, 2012

Copper coordination to the putative cell binding site of angiogenin: a DFT investigation. Luca Bertini, Maurizio Bruschi, Marco Romaniello, Giuseppe Zampella, Matteo Tiberti, Valentina Barbieri, Claudio Greco, Diego La Mendola, Raffaele P. Bonomo, Piercarlo Fantucci, Luca De Gioia[§], Theoretical Chemistry Accounts, 131, 1186, 2012

Mechanisms of intramolecular communication in a hyperthermophilic acylaminoacyl peptidase: a molecular dynamics investigation. Elena Papaleo[§], Giulia Renzetti, Matteo Tiberti, PLoS ONE, 7 (4), e35686, 2012

Molecular determinants of enzyme cold adaptation: comparative structural and computational studies of cold- and warm-adapted enzymes. Elena Papaleo[§], Matteo Tiberti, Gaetano Invernizzi, Marco Pasi, Valeria Ranzani, Current Protein and Peptide Science 2011, 12 (7), 657-683, 2011

ENCORE: Software for quantitative ensemble comparison. Matteo Tiberti, Elena Papaleo, Wouter Boomsma, Kresten Lindorff-Larsen[§], *To be submitted*

*: the authors equally contributed to this work

§: corresponding author

Chapter 1

Introduction

1.1 Metal atoms in protein systems

Many proteins coordinate one ore more metal atoms. It has been estimated that 40% of proteins require one or more metal ions to be able to carry out their biological function in cells [12, 96]. About twelve different metals are found in association with proteins in living systems, showing a great variety of protein-metal coordination modes, as well as biochemical and structural functions. Twelve metals, Na, K, Mg, Ca, Mn, Fe, Co, Ni, Cu, Zn, Mo and W appear as cations coordinated by one or more amino acid residues of the protein molecule or by protein-bound cofactors. Some, notably Mg, Fe, Co, Mo and W are often or always found as components of cofactors. There also are some additional borderline cases, such as Cr and V [12, 96]. Metals are the simplest, but most versatile, cofactors in protein biochemistry with a plethora of distinctive properties such as electron-acceptor ability, positive charge, flexible coordination sphere, specific ligand affinity, varying valence state, low- or high-spin configuration, and high mobility or diffusivity [13].

Metal-binding proteins are involved in a high number of cellular an physiological processes in life. Among the others, metal atoms have functional roles in intra-cellular and inter-cellular signaling, respiration, photosynthesis, oxygen transport, biosynthesis, electron transfer, biodegradation, drug metabolism, proteolysis and hydrolysis of amides and esters,

environmental carbon, sulfur and nitrogen cycles, and disease mechanisms [63].

Metal ions also often have a structural role in proteins. The most prominent example is probably calcium, whose binding is known as an important stabilizing feature in several proteins. Moreover, metal coordination can impose changes to protein structure and dynamics, which, in some cases, result in the modulation of function [47]. This mechanism is indeed exploited by the cellular machinery, in which Ca^{2+} acts as a physiological and cellular signal carrier. One of the most relevant examples of this behavior is the protein calmodulin, a calcium sensing protein in which calcium binding or unbinding affects its ability to bind to different protein targets, thus activating or deactivating different downstream signal cascades depending on calcium availability [3, 217].

Importantly, metal ions, especially those from the transition metal series, are involved in catalytic mechanisms of enzymes. Enzymes in which one or more metal atoms are essential for activity are thus collectively named metalloenzymes. More in general, enzymes are classified according to a rational system called Enzyme Commission (EC), in which a unique numeric identifier is assigned to each enzyme-catalyzed reaction. Enzymes are divided into six main classes (oxidoreductase, transferases, hydrolases, lyases, isomerases and ligases), and each of them is further divided into subclasses and sub-subclasses according to a hierarchical scheme [16]. Among the enzymes with known structure, at least 558 EC numbers include metal-dependent enzymes, spanning among all the six main EC classes. This number represents about 40% of all the enzymes with known structure in the Protein Data Bank (PDB), indicating the pervasiveness of metal ions in enzymes. Furthermore, metal-dependent enzymes occur in the 76% of all the subclasses covered by PDB, indicating that metal ions are involved in an extremely large variety of catalytic mechanisms [1, 11].

Metalloproteins usually coordinate metals through donor atoms in the functional groups of amino acid side chains. Of the 20 canonical aminoacids present in proteins, only a relatively small number are potential metal ligands. The ligand groups, which are encountered most often

are the thiolate of Cys, the imidazole of His, the carboxylates of Glu and Asp, and the phenolate of Tyr. Less frequently, metals can be coordinated by the thioether group of Met, the amino group of Lys and the guanidino group of Arg, and the amide groups of Asn and Gln. Metal ions can also bind to backbone atoms, through the carbonyl or the deprotonated amide nitrogen, and to the terminal amino and carboxyl groups of the protein. In many cases metal ions are not bound directly to the protein structure; instead, they are coordinated by a prosthetic group which is bound to the protein structure through covalent bonds or non-covalent interactions. This happens mostly with transition metals which are somehow involved in redox reactions. The most popular example of this case is that of the heme prosthetic group, which consists of a single functional iron atom coordinated by the heterocyclic organic ring of a porphyrin [63].

Of course, coordination geometries vary by atomic element, oxidation state and number of ligands. Not surprising, metallic ions preferentially bind different ligands according to their hardness or softness, i.e. to the respective degree of Lewis acidity. Hard Lewis acids such as K^+, Ca^{2+}, Mg^{2+} and Fe^{3+} bind preferentially to hard bases, such as oxygen atoms of carboxylates. This withstanding, even within this class, different elements feature varying coordination geometries due to differences in the specific charge-to-size ratio. For instance, Mg^{2+} and Ca^{2+}, which are very close in the periodic table, feature quite different coordination preferences. The Magnesium ion, which is significantly smaller than Calcium ion, is generally characterized by a strict octahedral coordination geometry, whereas calcium features an irregular coordination geometry, with sensible variations in the coordination number (7-10), bond length and angles [63]. The different physico-chemical properties of the two ions allowed evolution to assign different roles to them, since Calcium mostly has a structural role while the Magnesium ion is involved in catalysis, as detailed below [1].

In the case of ions with softer character, which usually are transition metals such as Cu^+ or Fe^{2+}, the directional covalent character of the coordination bond is prevalent and the ligands tend to dispose along the "ideal" coordination geometries for the given element, oxidation state and

spin multiplicity. Nonetheless, the protein structure can exert significant strain, thus imposing important and sometimes dramatical geometrical distortions. Depending on the single case, these softer ions are more often associate with S or N ligands, such as thiolates of Cys residues and the nitrogen atoms of the imidazole rings of histidine. Besides, several centers which include transition metals feature quite unique coordination environments, as for example exotic non-protein ligands such as inorganic sulfur atoms, CN^- or CO [10, 12, 63, 74].

While different metal species are generally more prone to catalyze specific reaction types, there is no strict one-to-one relationship between individual metals and functions, so that the same metal can carry out different functions or more than one metal can provide similar functions. This reflects the fact that metals have been selected for biological functions according to constraints other than mere functional characteristics, such as their relative abundance and availability in the environment . The choices taken by evolution thus reflect the adaptation to various conditions of metal bioavailability, which can change across space and has, sometimes dramatically, changed over time [1]

1.2 Activation of small molecules in metalloenzymes: relevance in a green economy world

Fossil fuels are the primary source of energy in the world, and their widespread adoption has been responsible for the increasingly high concentration of carbon dioxide in the atmosphere, which is connected to environmental concerns of primary importance, such as global warming and oceans acidification. Recent reports show that atmospheric CO_2 level has risen from about 278 ppm to 400 ppm since the industrial revolution, and it is expected to continue to rise [192]. Climate modeling studies show that the effects of climate change due to the increase in greenhouse gases would have lasting effects of up to one thousand years even when there is zero emission of greenhouse gases. In order to take on these and

other related issues, the implementation of a more sustainable economic and thus energetic model is essential. More in general, a green economy model has been defined "as one that results in improved human well-being and social equity, while significantly reducing environmental risks and ecological scarcities". In a green economy, the shift from hydrocarbons to more sustainable energy sources, such as the renewable ones, is seen as mandatory in the medium-long term. While fossil fuels are by large the mostly used energy provider, the use of renewable sources is indeed rapidly increasing together with the investment and involvement in national governments in this field. The use of non-fossil fuels and of energy source with intermittent and fluctuating output, however, requires the use of efficient energy storage systems and vectors [166].

Indeed, the simple dihydrogen molecule has been prominently proposed as a viable energy vector, because of its high energy density, and because its only oxidation product is water. Nonetheless, current hydrogen production mostly involves energetically costly processes, such as gas reforming from hydrocarbons or water electrolysis. To achieve a sustainable hydrogen economy, more efficient hydrogen production and storage technologies are required [107, 146].

Another promising route to chemically store renewable energy is the reduction of carbon molecules to obtain fuels which could be easily integrated in the already-existing energy infrastructure. An exceptionally appealing proposal in this context is the reduction of carbon dioxide to compounds such as methane, methanol, formate and hydrocarbons. This would allow at the same time to reduce the release of carbon dioxide in the atmosphere and to obtain a carbon-neutral fuel or C1 feedstock molecules that could be then exploited for further chemical processing [15, 85, 130, 132, 134, 214, 222].

Carbon dioxide reduction and sequestration is also involved in environmental remediation, which deals with the removal of pollution or contaminants from environmental media such as soil, sediment, or surface water for the general protection of human health and the environment. Indeed, the removal of pollutants from soil of water is a primary target in the context of green economy, since it would allow to rebuild limited eco-

logical resources and would help the transition towards a more sustainable agricultural system [166].

It is thus evident that key reactions involved in energy storage and transportation, as well as in general synthesis and environmental remediation, are of paramount importance for the implementation of a green energy infrastructure. Nonetheless, most chemical reactions are too slow and too unselective under normal conditions for direct industrial applications [72]. Reaction pathways can be greatly accelerated at high temperatures. Indeed, most industrial uncatalyzed synthetic processes require heating, introducing an extra energy cost and the production of significant quantities of waste (solvents, by-products...) that have to be reprocessed. The availability of effective, efficient and selective catalysts is thus fundamental for a green economy [21, 72]. To this extent, natural catalysts such enzymes are an interesting and appealing option. They have the capability of accelerating thousands of cellular reactions so that they are energy-efficient, highly chemo-, regio- and stereoselective, at both ambient temperature and pressure. Enzymes are routinely used in several industrial contexts, such as in the food, detergent, oleochemical and biomedical sectors [72].

In the context of green economy, metal-binding enzymes catalyze reactions that are able to activate small important molecules of crucial relevance. Enzymes that are involved in the redox decomposition, evolution or transformation of small molecules like H_2, CO, CO_2, CH_4, N_2, are believed to have first appeared in very ancient times, between 4.6 and 3.5 billion years ago.,when the atmosphere was probably rich in such gases and the oceans contained relatively high concentrations of metal species such as Fe^{2+} and Ni^{2+}. Indeed, enzymes such as hydrogenases, CO-dehydrogenases, nitrogenases, formate dehydrogenases, methyl–coenzyme M (CH_3-SCoM) reductase contain relatively exotic active sites in which one or more metal atoms are implied in the reaction [74].

Using purified enzymes in mass-production contexts can be however not very practical, as they require complex and expensive purification procedures and are quickly inactivated during catalysis [72]. It would be thus advisable to exploit the outstanding efficiency and specificity of

enzymes by reproducing their catalytic features in smaller, more stable, easily synthesized and manageable molecules. This approach is based on the principle that the increased understanding, at the very molecular level, of a given biological catalyst allows to reproduce its mechanicistic features in much simpler model complexes (biomimetic chemistry). Nonetheless, it is possible to explore unnatural combinations of atoms and molecules, that organisms have neither experienced nor kept during evolution for multiple reasons (bioavailability, selection pressure, toxicity...) to generate an original catalysts (bioinspired chemistry) for the same function. The variations introduced in the latter case are expected to improve the catalytic efficiency of the simplest biomimetic model and expand the range of substrates that can be transformed as well as the types of reactions catalyzed [33, 133].

Understanding the basic principles that drive the catalytic efficiency of metalloenzymes in terms of structure-function relationships is thus of paramount importance in order to drive the design of novel biomimetic catalysts and novel bio-inspired chemistry.

1.3 Structure-function relationships in metal-binding enzymes

1.3.1 Role of the metal atoms in catalysis

From a functional point of view, several other classification attempts have been carried out in the previous years, based on activity, bound metal and active-site structure, allowing to correlate structural features of the metal centers with the catalyzed reactions. In general, the catalytic mechanism of metalloenzymes can be classified into two macrocategories: the cases in which the metal atom aids catalysis by acting as a Lewis acid and those in which it is directly involved in a redox reaction (i.e. the metal's formal oxydation number changes) [11, 13].

In the first case, while a specific mechanism is of course different on a case specific basis, it includes the induced polarization of the bonds of a ligand, which may be the substrate or another molecule coordinated to

the metal atom. In the simplest case, the metal coordination increase the electrophilicity of the substrate, easing the attack by an electrophilic moiety. In other mechanisms of activation, metal binding results in increasing the acidity of a substrate/cofactor proton, which can be transferred within the catalytic site, promoting electronic rearrangement of the substrate/cofactor. Alternatively, proton release from the substrate/cofactor determines the formation of a reactive anionic species suitable for nucleophilic attack [11, 33, 47].

Moreover, the residual positive charge present on the metal ion can be used to counterbalance local negative charges formed in the active site during the course of the reaction. This effect can be additional respect to the Lewis acid destabilization of the enzyme-substrate complex outlined before, or be present by itself. This suggests that metals can provide a twofold contribution to lower the activation energy of the enzyme reaction, by both activating (i.e. increasing the energy of) the reacting species and stabilizing (i.e. decreasing the energy of) the transition state [1, 47, 49].

The metal ions that take part in these mechanisms are usually hard species with a high charge-to-radius ratio, such as those belonging to group II (most prominently Mg^{2+} and less often Ca^{2+}) together with a couple of d-block metal ions (Zn^{2+} and Mn^{2+}). Zn^{2+} in particular is widely represented in this category and represents a special case for being a d-block element, having its 3d shell completely filled. Zinc resembles the group II elements in that it is also stable in the +2 oxidation state, and is redox inert. Mn^{2+} has borderline softness characteristics, and it is able to carry out its function in some cases [1, 47, 49].

Things work differently when considering metal ions that are implied in redox reactions. Metal ions may be implied in either helping performing the reaction itself or help in transferring electrons to or from the active site. In this case, the considered species are transition metals having a softer character respect to those considered in the previous chapter, such as Fe^{2+}, Cu^{+} and Mo^{6+} [1, 47, 49].

In general the chemistry of these catalysts at the active site cannot be simply ascribed to a simple dominating effect, as happens with the non-

redox enzymes previously exposed. In fact, since in these cases metals are redox-active and have a fundamental role in the catalytic process, the specific stereoelectronic properties of the metal center are much more relevant in determining the reactivity. Since the d-shell electron of these softer ions are much polarizable than what happens with the outer shell electrons of harder species, especially the first coordination sphere and the active center geometry are of fundamental importance. Indeed, redox-active metal centers often feature peculiar first coordination spheres, which also include atoms from non-protein ligands or are altogether included in larger cofactors [1, 47, 49, 74]. Nonetheless, while the coordination geometry tends to be close to ideal one for the given metal atom and ligands, the protein sometimes poses great constraints on the active site geometry with sometimes dramatical distortions, as previously introduced. These distortions are thought to have a functional role, connected to the The state at which the metal center is found is in these cases is called an "entatic state", which has been defined as one closer to a transition state that to a conventional, stable molecule. This fine-tuning of the properties of bound metal centers thus allows to preorganize the active site towards reactivity, at the expense of the energy required to build and maintain the protein structure, which is of course necessary to mantain the active site at the entatic state [200].

Here follows a description of the principal coordination environments and structure-function relationships for metal-binding enzymes coordinating Molybdenum, Iron and Zinc metal ions, which are found in the enzymes that will be thoroughly discussed in the following chapters. While the following sections are not exhaustive in any way, they are meant to give an idea of the type of chemistry that can be expected from these metal centers as well as give to the reader a general idea of the great chemical flexibility and variability that can be achieved by having metal atoms at the active site.

1.3.2 Molybdenum

While it is relatively rare in the earth's crust, Mo is the most abundant transition metal in seawater. When we consider that the oceans are the

closest we get today to the primordial soup in which life first arose, it is not surprising that Mo has been widely incorporated into biological systems, and the only organisms which do not require Mo use W instead. The biological versatility of Mo and W result not only from their redox-activity, ranging through oxidation states VI to IV, but because the intermediate V valence state is also accessible, they can act as interfaces between one- and two-electron redox systems, which allows them to catalyze hydroxylation of carbon atoms using water as the ultimate source of oxygen [47].

Another factor which characterizes Mo (and W) enzymes is that, with the exception of bacterial nitrogenase, instead of having the metal itself directly coordinated to amino acid side chains of the protein, they contain a molybdenum pyranopterindithiolate cofactor (MoCo), which is the active component of their catalytic site. This cofactor coordinates the metal ion via a dithiolate side chain. The MoCo cofactor can exist in the fully oxidised (Mo(VI)) and fully reduced (Mo(IV)) forms, with some enzymes generating the (Mo(V)) form as a catalytic intermediate [47, 104, 105].

All molybdenum-containing enzymes other than nitrogenase (in which molybdenum is incorporated into a $[MoFe_7S_9]$ cluster of the active site) fall into three large families, as exemplified by the enzymes xanthine oxidase, sulfite oxidase, and DMSO reductase. They each have a characteristic active site structure and catalyze a particular type of reaction [104, 105].

The first, which is epitomized by xanthine oxidase from cow's milk, has an
LMo(VI)OS(OH) core in the oxidized state, with one equivalent of the pterin cofactor (designated L) coordinated to the metal. These enzymes typically catalyze the hydroxylation of carbon centers, with the exception of carbon monoxide (CO) dehydrogenase. The active sites of members of the xanthine oxidase family have an LMo(VI)OS(OH) structure with a square-pyramidal coordination geometry. The apical ligand is a Mo=O ligand, and the equatorial plane has two sulfurs from the enedithiolate side chain of the pyranopterin cofactor, a catalytically labile Mo–OH group, and most frequently a Mo=S [35, 104, 105].

The second family includes sulfite oxidase and nitrate reductases. Similarly as the xantine oxidase family, the (oxidized) metal center has a single equivalent of the pterin cofactor, as part of an $LMo(VI)O_2(SCys)$ core, with a cysteine ligand provided by the polypeptide. Members of this second family catalyze the transfer of an oxygen atom either to or from a lone pair of electrons on the substrate, respectively. Although the molybdenum centers of both the xanthine oxidase and the sulfite oxidase families possess a square-pyramidal coordination geometry with an apical Mo=O as well as three sulfurs and an oxygen in the equatorial plane [104, 105].

The final family is the most diverse structurally, although all members possess two (rather than just one) equivalent of the pyranopterin cofactor and have an $L_2Mo(VI)Y(X)$ trigonal prismatic coordination geometry. DMSO reductase itself has a catalytically labile Mo=O as Y and a serinate ligand as X completing the metal coordination sphere of oxidized enzyme. Other family members have cysteine (the bacterial Nap periplasmic nitrate reductases), selenocysteine (formate dehydrogenase H), -OH (arsenite oxidase), or aspartate (the NarGHI dissimilatory nitrate reductases) in place of the serine. Some enzymes have S or even Se in place of the Mo=O group. Unlike the oxygen atom transfer reactions catalyzed by members of the sulfite oxidase family, which cycle between dioxo Mo(VI) and monooxo Mo(IV) species, members of the DMSOR family cycle between monooxo Mo(VI) and desoxo Mo(IV) species [139].

1.3.3 Iron

Iron, is the fourth most abundant element of the earth's crust and, after aluminum, the second most abundant metal. It is implied in functions essential for life of complex organisms, such as oxygen trafficking, storage and respiration. Iron features high chemical flexibility with the possibility of existing in various oxidation states, the principal being II (d6) and III (d5), although a number of iron-dependent monooxygenases generate high-valent Fe(IV) or Fe(V) intermediates during their catalytic cycle [47].

Fe^{3+} is quite insoluble in water and significant concentrations of water-soluble Fe^{3+} species can be attained only by strong complex formation,

whereas in contrast, Fe^{2+} is extremely water soluble. Fe^{3+} is a “hard” Lewis acid and will therefore prefer "hard" oxygen ligands like phenolate and carboxylate rather than imidazole or thiolate. In contrast, Fe^{2+} has a softer character and can accommodate both the "hard" oxygen-based ligands and the "soft" ligands, such as those containing nitrogen and sulfur: examples are histidine, protoporphyrin, cysteine, and inorganic sulfur [47].

Given its mechanistic flexibility and its abundance, it’s not surprising that the ion is involved in a wide range of reactions and in a number of different coordination environments.

Probably the most widely iron coordination environment is that of heme binding proteins, in which iron is coordinated by a porphyrin cofactor. The peculiar properties of heme-bound iron allows it to perform as an efficient oxygen binding and redox center. Indeed, from a catalytic point of view, heme proteins are known for oxygen activation, in which a $P^{+}Fe(IV)$ intermediate is formed, which then oxidizes a substrate. Alternatively, heme proteins are used as electron transfer proteins as components of electron transfer chains, in which they interact with the other components and accept electrons from reduced donor molecules to transfer them to appropriate acceptors [47, 69].

Iron sulphur proteins contain 1Fe/4S, 2Fe/2S, 3Fe/4S, 4Fe/4S, and some higher clusters e.g. 6Fe/6S, in which all iron atoms are bound either to cysteine’s thiolate or sulphur atoms, with few exceptions. The geometry of these centers depends on the number of iron and sulfur atoms, being a single tetrahedral, rombic for 2Fe2/2S clusters, and cuboidal or cubane for 3Fe/4S and 4Fe/4s. These proteins act both as centers involved in electron transfer and as catalytic sites for a certain number of reactions. They are very well known, especially the cubane clusters, for their involvement in electron transport, being often embedded within redox enzymes where they act as molecular wires, delivering electrons one at a time between redox couples which are physically separated [47, 112]. Nonetheless, several iron-sulfur metalloenzymes also exist. In these cases, such as dehydratases one of the iron atoms is not coordinated to the protein and can coordinate a substrate. The role of metal can be either to

supply the right conditions for the reaction to occurr, without directly being oxidized or reduced. Otherwise, as happens in S-adenosylmethionine binding enzymes, a reduced cluster can provide an electron thus generating a radical in the substrate which promptly reacts [47].

There are many other proteins that contain iron in a form which is neither in haem nor in iron-sulfur clusters. These sites bind one or two iron ions, or sometimes one of them together with another metal. Many of them are activate molecular oxygen for the oxidation of a target substrate. Although these enzymes catalyse a diverse range of chemical reactions, structural studies show that they share a common structural motif in their catalytic centre. This motif is made of one aspartate or glutamate residue, which coordinate the iron in a facial triad arranged at the vertices of one triangular face of an octahedron. In this way only one face of the coordination sphere, leaving three coordination sites occupied by displaceable water molecules, which can be replaced by exogenous ligands, such as oxygen, substrate, and/or cofactor. This gives the protein the flexibility with which to tune the reactivity of its Fe(II) centre, giving rise to the extraordinary range of catalytic versatility of these enzymes. Fe(III) mononuclear binding enzymes, although less prevalent, have also been identified [122, 191].

Dinuclear non-heme enzymes also exist, and their common functional trait is that they react with dioxygen as part of their function, and function as redox enzymes. They all contain a four-helix- bundle-protein fold, surrounding a di-iron core with the two iron atoms in close proximity, one or more bridging carboxylate ligands, with terminal carboxylate and/or histidine ligands, and often a bridging oxo, hydroxo, or aqua ligand, at least at the di-Fe(III) oxidation level [122, 190, 191].

Among the binuclear enzymes, hydrogenases are a somewhat special case, because of the peculiarity of their active site. From a structural point of view, these enzymes have been divided in three classes. Two of them are binuclear enzymes, one which binds two iron atoms ([FeFe]-hydrogenases), and one which has a iron-nickel center [NiFe]-hydrogenases. These enzymes catalyze the evolution of molecular hydrogen from protons and electrons alone, and the opposite reaction. The metal atoms have unique

coordination environments, with the iron atoms ligated by small inorganic ligands (CO and CN^-) and containing sulfur bridged bimetallic centers, typically with an open coordination site on one metal. A subgroup of the [NiFe]-hydrogenases is formed by the [NiFeSe] hydrogenase, in which one of the cysteine ligands of the nickel is replaced by a selenocysteine [75, 185, 204]. Finally, an hydrogenase binding a single iron atom ([Fe]-hydrogenases) exist, which catalyzes the reversible reduction of the substrate methenyltetrahydromethanopterin (methenyl$-H_4MPT^+$) with H_2 to methylene$-H_4$MPT and a proton, and doesn't contain any iron-sulfur cluster differently from the bimetallic hydrogenases. In this case, iron is coordinated by a nucleotidic moiety, forming the Fe-guanylylpyrodinol cofactor [108, 147].

1.3.4 Zinc

The divalent zinc ion (Zn^{2+}) is one of the most frequently found metal cofactor in enzymes, and it is present in more than 300 of them, spanning all the six main EC classes [10, 47].

It is very peculiar in the transition metal series because its Zn^{2+} ion has a completely filled d shell. This means that it is redox inactive, in contrast, for example, to manganese, iron, and copper. Given its relatively small ionic radius (0.65 Å), it has a highly concentrated charge and is one of the best Lewis acids among the metal atoms, being outclassed only by Cu^{2+}. From a structural point of view, its d10 configuration means that its complexes are not subject to ligand field stabilization effects, meaning that Zn^{2+} has no ligand field constraints on its coordination geometry. Coordination number and geometry are therefore dictated only by ligand size and charge, so that, in principle, Zinc centers can adopt an highly flexible coordination geometry. Nonetheless, in most zinc proteins, there is a strong preference for tetrahedral coordination, frequently distorted. In catalytic zinc sites the ion is usually coordinated with any three nitrogen, oxygen or sulfur donors of histidine, glutamate, aspartate and cysteine, with histidine being the most frequent amino-acid, usually through its Nϵ nitrogen. Zinc can be found in mononuclear or polinuclear centers, in which can be coupled to other Zn(II) as well as other elements

[10, 45, 47, 96, 185].

The reactivity of zinc in active sites is not surprisingly determined by its Lewis acidity. Nonetheless, other properties of this ion favor its use in enzymes as the catalytic moiety. Zinc thus features a flexible coordination chemistry, allowing to bind to a vast number of different ligand that are able to fine-tune its properties. On the other hand, the zero ligand field stabilization energy allows to rapidly exchange from one coordination geometry to another and fast ligand exchange, both appealing features from the point of view of catalysis [10, 18, 45, 47].

The catalytic mechanism of most mono-Zinc enzymes typically relies on central catalytically active Zn^{2+} atom bound to three protein ligands, and the fourth site occupied by a water molecule. The zinc-bound water is a critical component for a catalytic zinc site, because the coordination to Zn(II) greatly increases its acidity. The bound water can then be promptly deprotonated, obtaining a hydroxide ion in a controlled environment which readily reacts with other species through a nucleophilic attack. Otherwise, the bound water can be replaced by other moieties, especially of an oxygen atom from a C-O or P-O group, whose bond becomes polarized and prompted for a nucleophilic attack [10, 18, 47, 185].

In active sites hosting two or more metal atoms, whether they are all Zn(II) or Zn(II) with some other metal, the metal atoms are found to be in close proximity in one another, and may operate in concert to enhance catalysis. In these cases, when the catalysis is actually performed by the Zn(II) atoms, one of the zinc ion, designated as the catalytic zinc, coordinates a water molecule that is essential for catalysis and has coordination environment similar to those found in other single catalytic zinc sites. The remaining one or two metal ions may be involved in substrate coordination or transition state stabilization [10, 18, 47, 185].

1.4 Role of dynamics in enzyme function

In metalloenzymes the chemical step of catalysis is relegated to the metal-binding active center. While the general mechanisms by which enzymes achieve their fast reaction rates is still debated, it is undoubted that

the metal cofactor have an essential role in the catalysis mediated by metalloenzymes. As far as the catalytic step is regarded, both in metalloenzymes and others, the enzyme structure is thought to provide a pre-organized environment that induces a stronger stabilization of the transition state respect to the reactant state (which can be the bound enzyme-substrate complex or the free enzyme and substrate). Such a preorganized active site can lead to contributions to catalysis from interactions with the bound substrate or with the bound product of one step of a reaction that serves as the reactant in the next step. If the difference between the reactant state and transition state involves substantial charge transfer, as it does in many enzymatic reactions, the positions of the polar and charged groups in the enzyme (including metal ions and cofactors, as in metalloenzymes) play the essential role. Hydrogen bonding is often used in combination with other electrostatic effects to stabilize the transition state [86].

Nonetheless, the enzyme-catalyzed reaction is composed of many steps, with the chemical step being one of them. The turnover number from Michaelis-Menten kinetics model, k_{cat}, is the rate by which the enzyme converts substrate into product, including every necessary step from the ES complex (substrate-bound enzyme) to the dissociated product. Enzyme catalysis thus includes at least processes related to chemistry, often with several intermediate steps, and product release [206].

From a more formal point of view, the catalytic efficiency of a given catalyst (hence of enzymes as well) can be modeled using the energetic span model. This model, which has been designed to evaluate the TOFs (turn-over frequency) of the catalysts in a straightforward manner from theoretical free-energy profiles, shows that the overall apparent activation energy of the cycle, which governs the TOF, in most cases depends only on the free energy of one transition state and one intermediate of the whole cycle: the TOF-determining transition state (TDTS) and the TOF-determining intermediate (TDI). These are the states that exert the highest degree of control on the cycle itself, and are not required to be subsequent one of the other because of the cyclicity of catalysis. Once again, this enforces the notion that not only the catalytic step but also

others should be considered [124].

In this context, the hypothesis that protein motions are involved in enzymatic turnover has gained significant attention. Indeed, it is well known that proteins are not static entities but rather are intrinsically flexible molecules. Protein conformational fluctuations driven by thermodynamical motions enable the protein structure to sample a range of conformations and substates. Internal protein motions span time scales ranging over more than twelve orders of magnitude, including bond vibrations on the 10–100 fs timescale, rotations of side chains at the protein surface on the 10–100 ps timescale, hinge bending at domain interfaces on the 100 ps to 10 ns timescale, water structure reorganization on the 10 ns timescale, and rotation of medium-sized side chains in the protein interior on the micro- to millisecond timescale. Motions at fast time scales enable sampling of the conformations within the substates, while conformational transitions at long time scales enable the protein to transition from one substate to another [98, 170].

Consensus has grown around the concept that protein conformational changes take place during different steps along the turnover cycle of enzymes, whose frequency is in the millisecond timescale. Indeed, there are some evidences that the catalytic cycle of several enzymes may be controlled by conformational changes on the millisecond timescale, so that the protein structure occasionally samples high-energy intermediate states that are preorganized towards the crossing of the chemical barrier. The protein is thus stochastically expected to follow pathways encoded in the structure to optimize a catalytically competent enzyme–substrate complex [60, 94, 145].

This has been shown by both experimental and theoretical work or integration of them. For instance, in dihydrofolate reductase (DHFR) the participation of excited-state conformations, detected by NMR experiments, showed that the structural and dynamic changes that occurred in the protein in response to the binding and dissociation of substrates, cofactors and products were an integral part of the mechanism of the enzyme [28, 94]. In Cyclophilin A (peptidylprolyl isomerase A) the characteristic enzyme motions detected during catalysis are preserved in the free en-

zyme with frequencies corresponding to the catalytic turnover rates [66]. In these two cases at least, the enzyme conformations compatible with those of the reaction cycle. Mutations in the regions involved in conformational changes severely affect the reaction rate, and in the characterized instances, there is a concomitant change in the rate of conformational changes [94]. Similar behaviour has also been reported, for instance, in the metalloenzyme phosphotriesterase, as will be detailed in the following chapters.

While these steps are necessary to achieve the conformational of a functionally active enzyme, there is still debate on whether specific protein dynamics have a direct role (i.e. are coupled) with the chemical step.

A proposal in order to solve this issue is that is that fast femtosecond to picosecond dynamic motions at the transition state contribute significantly to catalysis. From this perspective, dynamics directly couple a promoting vibration to a transition state reactive bond. In this interpretation of enzymatic catalysis, the transition-state barrier is overcome by the formation of rare promoting vibrations at the catalytic site that lower the barrier for bond change. While of course the free energy barrier for the chemical step is related to the relative probability of sampling transition state configurations relative to reactant configurations, it's not clear whether the direct coupling of specific protein dynamical modes to the fast (femtosecond) energy barrier crossing is catalytically relevant in the sense of influencing the overall rate of the enzyme-catalyzed chemical reaction, and no conclusive proof exist of this direct coupling [184]

In another perspective, the active site of the enzyme simply is preorganized so that the reorganization of the environment required to evolve from the reactant to the transition state is lower in the enzyme than in solution, and the protein dynamics are not strictly coupled to the chemical step. Nonetheless, achieving the preorganized active site for reactivity (i.e. in contrast to when the active site is preorganized for substrate binding or product release) may require rate-limiting conformational changes within the turnover cycle, as previously detailed [114, 165].

Chapter 2

Aim

The present work aims to investigate structure-function relationships in metalloenzymes. In particular, we have addressed our attention to three systems of interest: Formate dehydrogenase (FDH), [NiFe]-hydrogenase and phosphotriesterase (PTE). As will be more clear in the following chapters, the former two enzymes both catalyze redox reactions through a metal center, in which a metal atom changes formal oxidation state upon substrate reduction or oxidation. Both of them are exceptionally interesting targets, both for the exceptional relevance of the catalyzed reactions and in the peculiarity of their active sites and reaction mechanisms, especially in relation with the chemistry requirements of a green economy society. Formate dehydrogenase catalyzes the conversion between carbon dioxide and formate through a Mo center, while the hydrogenase is involved in the evolution or decomposition of molecular hydrogen via a bimetallic Ni-Fe active site. As detailed in the introduction, both hydrogen and formate are especially relevant compounds to be used as energy vectors in a green economy world, and their catalysis is a prominent area of research. Understanding the mechanism of their molecular functioning would allow to design and test enzyme-inspired compound models and related biomimetic chemistry, in order to develope efficient catalysts for these reactions.

Phosphotriesterase are enzymes hosting a two-metal active site that hydrolize organophosphates and related compounds. This class of com-

pounds, which comprehends pesticides and nervine gases, represents one of the most diffuse pollutants in the modern world and a persistent threat for human health. Indeed, in the context of bioremediation, phosphotriesterases have been proposed as a viable catalyst able to decompose organophosphate pesticides efficiently and in a safe and enviro-friendly manner. While the reaction mechanism of phosphotriesterase is not completely understood yet, it is well known how protein dynamics play an important role in the reaction mechanism itself. In fact, at least for a number of substrates, it has been proposed that the slowest step of the catalytic cycle is not substrate hydrolysis per se, but involves a conformational equilibrium between different conformational basins in the active site. Aiming to understand the function-structure relationships in PTE, we have used classical molecular dynamics to investigate the dynamical properties of PTE wild-type and mutant variants, which concomitantly showed altered catalytic properties and differences in the conformational landscape relevant to the reaction.

Chapter 3

Formate oxidation in Mo-binding formate dehydrogenase

3.1 Introduction

Bacterial formate dehydrogenases are metalloenzymes able to catalyze the direct interconversion $HCO_3^- \rightleftharpoons CO_2 + H^+ + 2\,e^-$. In enterobacteria, such as *Escherichia coli*, they are involved in maintaining the correct redox balance and pH, by oxidizing formate to CO_2 and using the obtained electrons to reduce other species. They are indeed part of intermembrane complexes, which feature several cubane clusters that carry electrons to the active site of another enzyme, where reduction of a substrate takes place. Several variants of these complexes exist, binding different FDHs which carry out the reduction of different species and are expressed according to different cellular conditions. For instance, three different isoforms, FDH-H, FDH-N, FDH-O are coupled with, respectively, hydrogen evolution, nitrate reduction and ubiquinone reduction. In other more exotic species, such as *Syntrophobacter fumaroxidans*, the W-binding formate dehydrogenases uses the reducing equivalents coming from the reaction propionate $\longrightarrow$ acetate $+ CO_2 + 12\,e^-$ to produce hydrogen and reduce CO_2 to formate.

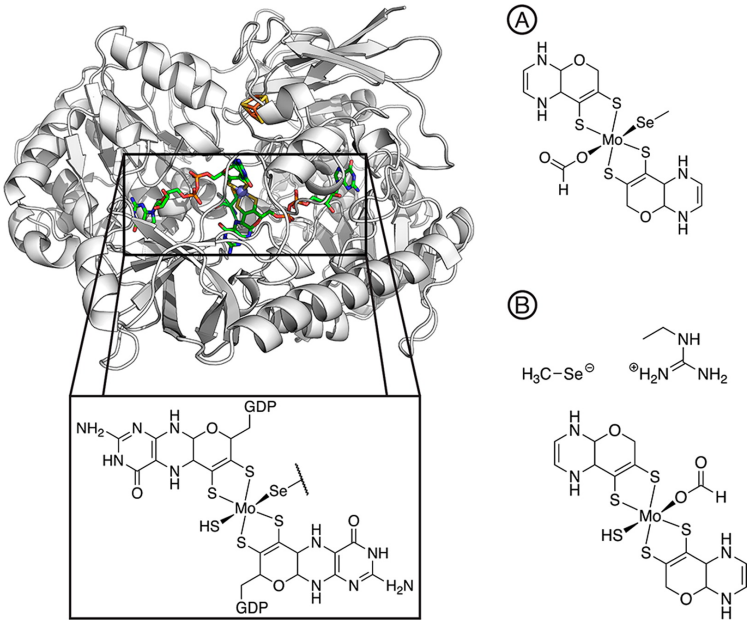

Fig. 3.1 Left: structure of FDH-H from *E.coli*, with scheme of the full Mo-coordinated bis(MGD), Selenocysteine selenate and the SH^- moiety. Right: Employed models (A and B).

Mo-binding FDHs belong to the dymethil sulfoxide reductase family, in which metal coordination takes place via bis-chelation of the S atoms of dithiolenes of two independent MGDs. In the FDHs, the coordination sphere of Mo(VI) is completed by the selenate of a deprotonated selenocysteine (Se-Cys) residue and a sulfur ligand (SH^- or S^{2-}). FDHs usually also host other metal centers, such as Fe_xS_y clusters, used for electron transport.

The reaction mechanism underlying the catalytic oxidation of formate by FDH has not been completely elucidated yet, although some proposals, deriving from both experimental and computational studies, have been reported [32, 129, 144, 167]. Unlike other members of the DMSO reductase family, the FDH catalytic reaction involves direct proton transfer,[118] which is the rate-determining step of the whole catalytic process [144]. Previously proposed mechanisms for formate oxidation start with the replacement of one of the Mo^{VI} ligands by a substrate, with subsequent

proton abstraction and electron transfer to the metal ion, which is thus reduced to Mo^{IV}. Two models of the reaction mechanism have been suggested for the proton-transfer step, both involving the same proton acceptor: in the first (usually denoted as A), formate displaces the sulfur ligand (SH^- or S^{2-}) and proton abstraction is performed by the Se atom of Se-Cys, which remains coordinated to the Mo ion [32]; in the second (B), formate replaces the selenium ligand in the metal coordination environment, and proton abstraction is performed by the unbound selenide [167]. Such two reaction pathways have been previously investigated by density functional theory (DFT), with mechanism B as both thermodynamically and kinetically favored [129]. A more recent work [144] on the modeling of the full catalytic cycle also proposed that the proton abstraction step is performed by the unbound Se-Cys. Finally, the initial part of the reaction mechanism has been recently investigated, proposing a common mechanism for substrate coordinate to the metal atom in some Mo-binding enzymes of the DMSO reductase family, such as FDH and periplasmic nitrate reductase.

3.2 Aim

Even though the basic properties of Se-Cys and its location within the active site are consistent with the proposed role of the proton acceptor in the FDH catalytic cycle, nevertheless, other atoms coordinating the Mo^{VI} ion can act as putative proton acceptors: (i) the S atoms belonging to the dithiolene moiety of the MGD cofactor, (ii) the SH^- (or S^{2-}) group, and (iii) the Mo ion itself, through a hydride-transfer step. Therefore, with the aim of shedding further light on one of the key steps in the reaction catalyzed by FDH, we have investigated by DFT all of the possible pathways of H^+ or H^- transfer from formate to the enzyme active site.

3.3 Methods

Density functional theory (DFT) has been used in order to calculate energies and geometries of the stationary points of the studied reactions. All

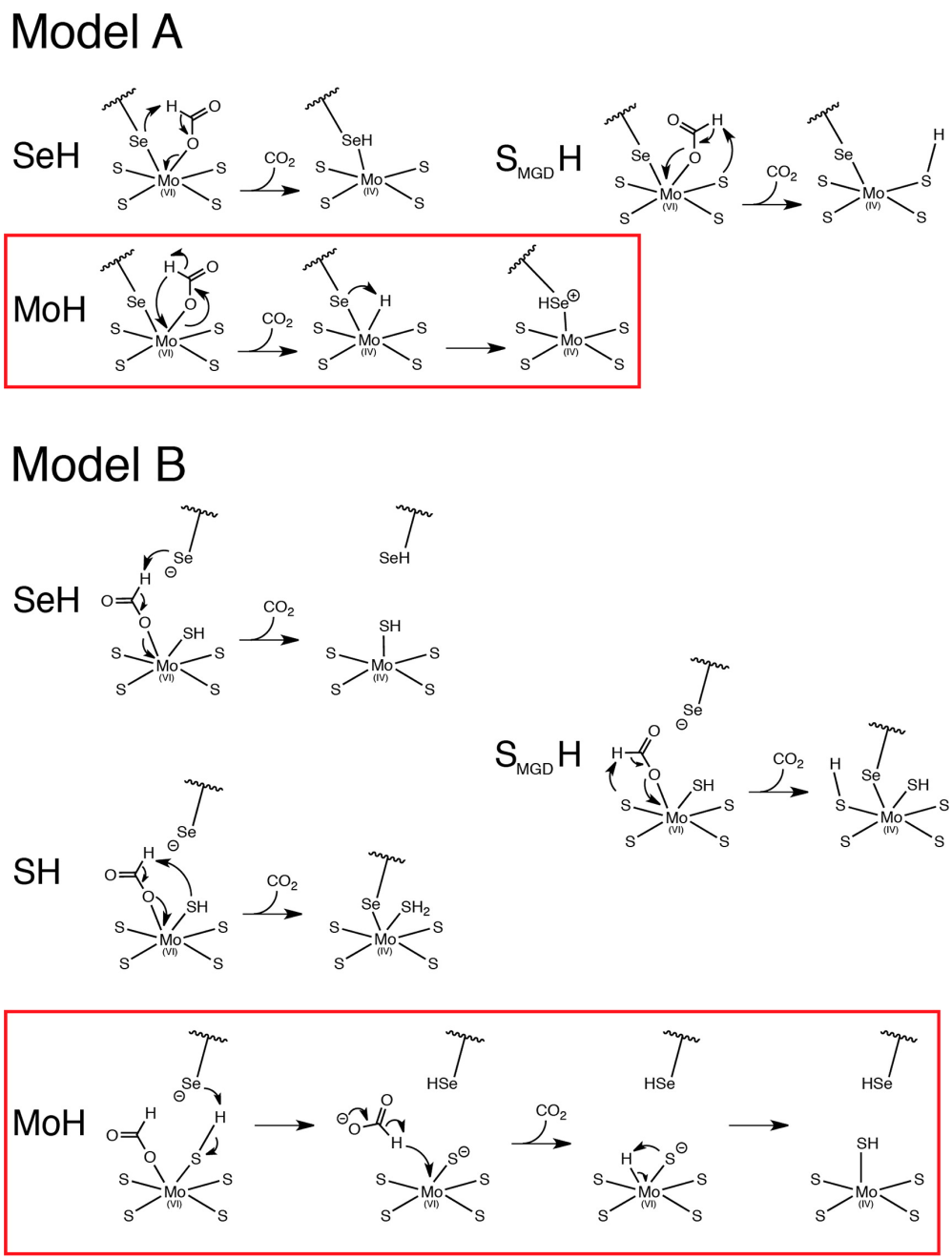

Fig. 3.2 Investigated Reaction Mechanisms of H Transfer Related to FDH Catalysisa. The only transfer occurring as a hydride (via a metal center) is boxed in red.

of the performed DFT calculations have been carried out using a valence triple-Z basis set with polarization on all atoms (TZVP) [181] and the pure exchange-correlation functional BP86. The inner-core 28 electrons of the Mo ion have been treated using a relativistic effective-core potential (ecp-28-mwb; λ max = 3) [9], whereas the valence electronic structure has been investigated by the same full-electron basis (TZVP) that has been employed for the rest of the system. The chosen computational setup permitted to accurately reproduce the distances between the Mo^{VI} ion and the S and Se atoms of the first coordination sphere, as found in the *E. coli* FDH-H crystal structure (PDB ID 1FDO [32]). Moreover, the BP86 functional has been shown to appropriately describe chemical systems involving CO_2-metal interactions [83, 109]. The reported calculations have been performed by employing a protein-like dielectric constant value (ε = 4) as implemented in the COSMO [182] solvation model. Anyway, it has been verified that moderate changes (less than 2 kcal mol^{-1}) can be detected by switching from energetics computed in vacuo and by including an implicit solvation model. DFT calculations have been carried out using the TURBOMOLE suite of programs [4]. Stationary points of the energy hypersurface have been located by means of energy-gradient techniques. Geometry optimization has been carried out on starting structure guesses. Starting atomic positions were derived, for both models A and B, from the *Escherichia coli* FDH-H crystal structure (PDB ID 1FDO), in which the Mo atom was in the oxidized state of Mo^{VI}. For model A, the SH^-ligand was replaced by a formate molecule. For model B, the CH_3Se^- moiety was moved away from the Mo coordination sphere, interacting with the arginine (Arg) moiety. A formate molecule was placed instead in the Mo coordination sphere. Amino acid models employed in the present study were permitted to move freely during the optimizations, in order to avoid obtaining artificial negative eigenvalues in the Hessian matrix evaluation. Indeed, the actual shift of such residues that we used for B-type mechanisms has been observed to be quite limited in extent, and especially it has been found to be compatible with the overall structure of the enzyme active-site region, as observed in the crystal structures. Moreover, it is noteworthy that the region that hosts the active site can

be highly dynamical: for instance, the Se-Cys residue has been found to be able to move drastically (up to more than 12 Å) with respect to its position in the Mo-bound form [167]. We thus deemed that fixing residues or tightly constraining them in their initial positions would have limited their conformational freedom in an unrealistic fashion.

The optimization of transition-state structures has been carried out according to a procedure based on a pseudo-Newton-Raphson method. Initially, geometry optimization of a transition-state guess geometry is performed, keeping the distances corresponding to the selected reaction coordinate constrained. Normal-mode analysis at the BP86/TZVP level is then carried out on the constrained minimum-energy structures, and if one imaginary vibrational frequency corresponding to the reaction coordinate is found, the curvature determined at such a point is used as the starting point for the next phase. Finally, the transition-state search is performed according to an eigenvector-following search procedure, in which the eigenvectors in the Hessian are sorted in ascending order, with the first one being that associated with a 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. NBO/NPA [172] partition schemes of the electron density have been used for determining the atomic partial charges. All of the ΔE values discussed in the present contribution have been calculated as differences between pure electronic energies. A B3LYP [24, 25, 128, 194] exchange-correlation functional has finally been used to check the consistency of the BP86 results.

3.4 Results

We have studied the proton-transfer step using the two models of the active site, as detailed in the Introduction, that were adopted in precedent DFT studies: in the first, formate replaces the SH^- group in the Mo coordination sphere (A; Figure 3.1, top right); in the second, formate displaces the Se ligand (B; Figure 3.1, bottom right).

Referring to the aforamentioned models, the two different cofactor

configurations will be referred to as mechanism-model A (denoted as MA) and mechanism-model B (denoted as MB). In both cases, the enzyme active site has been modeled including the Mo center and some of the surrounding residues, which are found in the E. coli FDH-H. Previous DFT studies [129, 144] of such systems used two simple dimethyldithiolene moieties to model the MGD cofactors; however, a very recent investigation pointed out that the organic portion of the cofactor (and its synthetic analogues) is non-innocent with respect to redox reactions and that a more extended model should be used to better represent the stereoelectronic properties of the MPT [138, 177]. Therefore, the two MGD groups have been represented as dithiolenepyranopyrazine, which has been reported to be the optimal candidate in terms of the accuracy/computational cost ratio [177]. The inorganic S ligand coordinated to Mo was modeled initially as SH^- [129], and the Se-Cys residue was modeled by CH_3Se^- (or by the full residue, *vide infra*). Because the MB model features a more negative overall charge with respect to MA, due to the presence of both CH_3Se^- and SH^-, a protonated dimethylguanidine group was added to mimic Arg333 in order to offset the negative charge. The following proton/hydride acceptors were tested for H-CO_2 transfer in both models A and B (see figure 3.2 for formal reaction schemes): i) the Mo-bound Se atom (referred to as SeH-A mechanism in the following) and the Mo-unbound Se (SeH-B mechanism); ii) one of the MGD thiolenes (S_{MGD}H-A and -B mechanisms; with Se-Cys either bound or unbound to Mo, respectively); iii) the Mo(VI) ion, through hydride transfer (MoH-A and -B mechanisms; with Se-Cys either bound or unbound to Mo, respectively). In pathway iii) also the subsequent H^+ transfer step from the metal hydride to the Se or S atoms found in the active site has been investigated.

Concerning MB, the case in which the SH^- ligand performs proton abstraction has also been considered, which is not possible in MA (where SH^- is absent because it is replaced by formate). In addition, previous studies illustrating only SeH pathways pointed out that the presence of a conserved histidine (His) residue substantially lowers the energy barrier for proton transfer to the unbound selenide. To take this effect into account, the SeH-B reaction pathway has been studied using a modified

MB in which the entire His-Se-Cys dipeptide (the two residues are consecutive in the sequence of FDH-H) was included. The reaction pathways have been investigated adopting the S = 0 spin state for MA and the S = 1 spin state for MB. Such different spin multiplicity characterizing MA and MB has been recently determined by Russo and co-workers, in the case of Se-based H^+-transfer pathways [129].

The proton-transfer reaction involving S or Se as a final acceptor occurs as an elementary reaction event, whereas the MoH-A and -B pathways are characterized by the formation of an intermediate species. In both MA and MB, H^+ transfer to the MGD thiolenes always features high energy barriers (33.6 kcal mol^{-1} for MA; 44.8 kcal mol^{-1} for MB), and this holds true also for transfer to the SH ligand in MB (41.2 kcal mol^{-1}). MoH-A and SeH-A pathways are characterized by energy barriers that slightly favor the former (15.9 and 18.6 kcal mol^{-1}, respectively), while in the SeH-B pathway, the energy barrier is significantly higher than that in MoH-B (22.5 and 11.5 kcal mol^{-1}, respectively). As for thermodynamics, all MA pathways are endergonic, whereas among MB pathways, the SeH, $S_{MGD}H$, and MoH routes feature exergonic profiles.

Structural analysis of the stationary points located using DFT shows that proton transfer to either Se or S involves a simple proton shift from formate to the acceptor atom. In the reactants (in both MA and MB), the metal center has a trigonal-prismatic coordination geometry, with one formate O atom and a Se or S atom (depending on the model) to complete the coordination environment of Mo. In MA, the Se ligand is coordinated to the Mo ion, whereas in MB, it interacts with Arg (2.43 Å). Such an interaction is consistent with the crystal structure of reduced *Escherichia coli* FDH-H, in which the uncoordinated Se-Cys was found to interact with the corresponding R333 residue [167]. Arg also forms a hydrogen bond with an O atom of the coordinated formate (1.72 Å).

In the $S_{MGD}H$ mechanism, the H^+ acceptors are the MGD S^- atoms. Only proton transfer to one of the four available S atoms has been studied for symmetry reasons. Analysis of the $S_{MGD}H$-A pathway highlights the rotation of the formate molecule and the concomitant slight bending of the Mo-$O_{formate}$-$C_{formate}$ angle going from the reactant to the transition

state. This structural reorganization allows the formate H atom to come closer to the acceptor S_{MGD} atom. The $C_{formate}$-H bond is elongated to 1.78 Å in the transition state, while the S_{MGD}-H distance is 1.46 Å. The reaction product features a square-pyramidal geometry of the Mo center, while CO_2 moves away from the active site. As for S_{MGD}H-B, the proton-transfer reaction pathway is similar to that of S_{MGD}H-A, with $C_{formate}$-H and S_{MGD}-H distances of 1.49 and 1.95 Å, respectively. In the products, the Se ligand, which was not coordinated to the metal ion in the reactants, binds to Mo, which attains a slightly distorted octahedral coordination geometry.

The SH reaction pathway can take place only in MB reaction models. The mechanistic and structural features of the transition state are analogous to those previously discussed for S_{MGD}H. The Arg side chain interacts with both formate and the selenide of Se-Cys. The transition state for H^+ transfer is characterized by a $C_{formate}$-H distance of 1.46 Å and a S-H distance of 1.97 Å. Also, in this case in the product, the CH_3Se^- group is coordinated to the Mo ion, forming a distorted octahedral geometry. In addition, CO_2 is displaced from the active site and the guanidinium group interacts with S_{MGD} atoms and the Se atom.

The hydride transfer from formate to molybdenum (MoH) has been investigated for both MA and MB. Notably, it turned out that this pathway features low energy barriers. Hydride transfer in MoH-A occurs according to a two-step mechanism, i.e., through an intermediate (IA) formation, which connects the two transition states TS1A and TS2A. In TS1A, the Mo complex retains the overall distorted trigonal-prismatic coordination geometry encountered in reactants, while the coordinated formate tilts over the metal ion so that the H atom gets closer to Mo (2.02 Å) and the $C_{formate}$-H bond is stretched from 1.12 Å in the reactant up to 1.36 Å. IA is then formed, which features the $C_{formate}$-H bond shortening to 1.22 Å, whereas the Mo-H and Mo-$O_{formate}$ distances are almost unchanged compared to TS1A. In addition, IA shows a significant elongation of the Mo-O bond compared to the reactants. TS2A is associated with the actual hydride transfer to Mo: the $C_{formate}$-H distance increases to 1.62 Å, and also the $O_{formate}$-Mo distance increases to 2.77 Å. Such a step can be

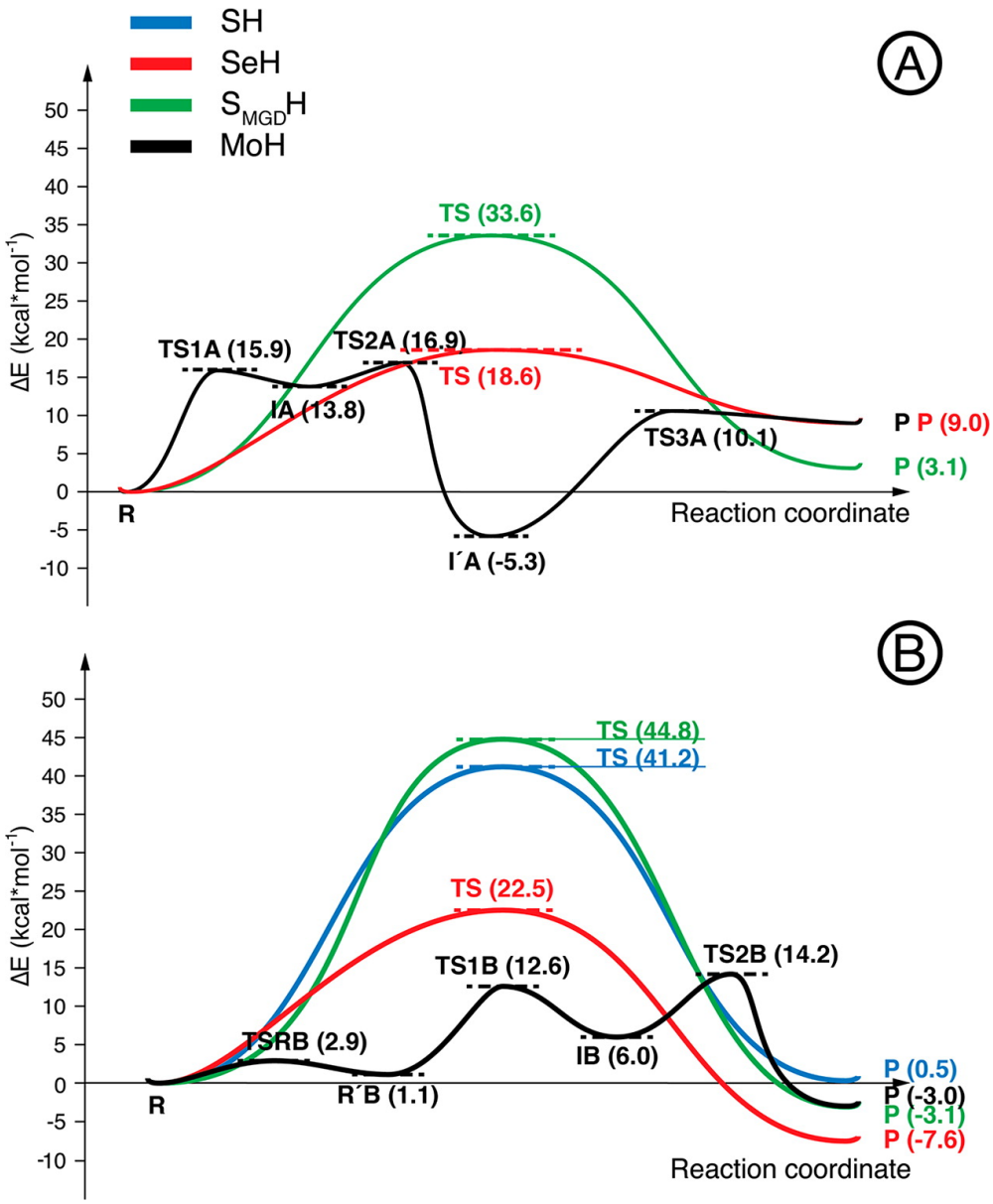

Fig. 3.3 Reaction profiles associated with all investigated pathways of proton transfer at the DFT/BP86 level.

referred to as an actual β-elimination of H, *vide infra*). The Mo-H product (I'A) shows a distorted trigonal-prismatic coordination of the Mo center, with CO_2 leaving the active site. Finally, proton transfer from Mo to the coordinated Se ligand closes the $H^+/2e^-$ process. The barrier associated with this step (see TS3A; Figure 3.3) is 15.4 kcal mol^{-1}. The energetic profile of the entire MoH-A pathway reveals the key effect exerted by the Mo ion, which acts as a mediator of the H-transfer process.

The MoH-B pathway is quite different from the corresponding MA route. It is worth recalling that in MB the Se ligand is not originally bound to Mo. In fact, in the reactant, the Se atom is replaced by one of the formate O atoms. Nevertheless, as observed in the previous $S_{MGD}H$ and SH mechanisms, the Se coordination to Mo is restored because of the presence of a vacant site on Mo, which opens up as the produced CO_2 detaches from the metal ion. Differing from the other two mechanisms, the MoH reaction pathway entails the formation of metal hydride as a product, thus saturating the first coordination sphere of Mo. Several reaction pathways have been tested for the MoH-B reaction mechanism. Remarkably, the one characterized by the lowest-energy barrier features SH deprotonation and concomitant CH_3Se^-protonation (Figure 3.4; (R−>R˙). Mota et al. indeed proposed that the SH ligand could indeed be found in an unprotonated form [144]. Moving from R', hydride transfer and simultaneous decarboxylation occur through TS1B (Figures 3.7 and 3.8) in a single step, thus differing from MoH-A.

TS1B features the decoordination of formate from Mo ($O_{formate}$-Mo: 3.88 Å), while in the reactant geometry associated with this catalytic step, the formate O atom is still coordinated to Mo. Once again (see MoH-A description), when referring to such a peculiar reaction intermediate (R'B, Figure 3.4), the proposed mechanism can be conceived to involve β-elimination of a H atom to the metal ion and is typical of metal formates that undergo decarboxylation, forming the corresponding hydrides. Such a crucial event is also occurring during catalysis of the water-gas shift reaction [52, 53, 65]. It is worth noting that whereas in MoH-A the term β-elimination can be used both referring to reactants and their subsequent TS, in MoH-B such definition can be adopted only when referring to

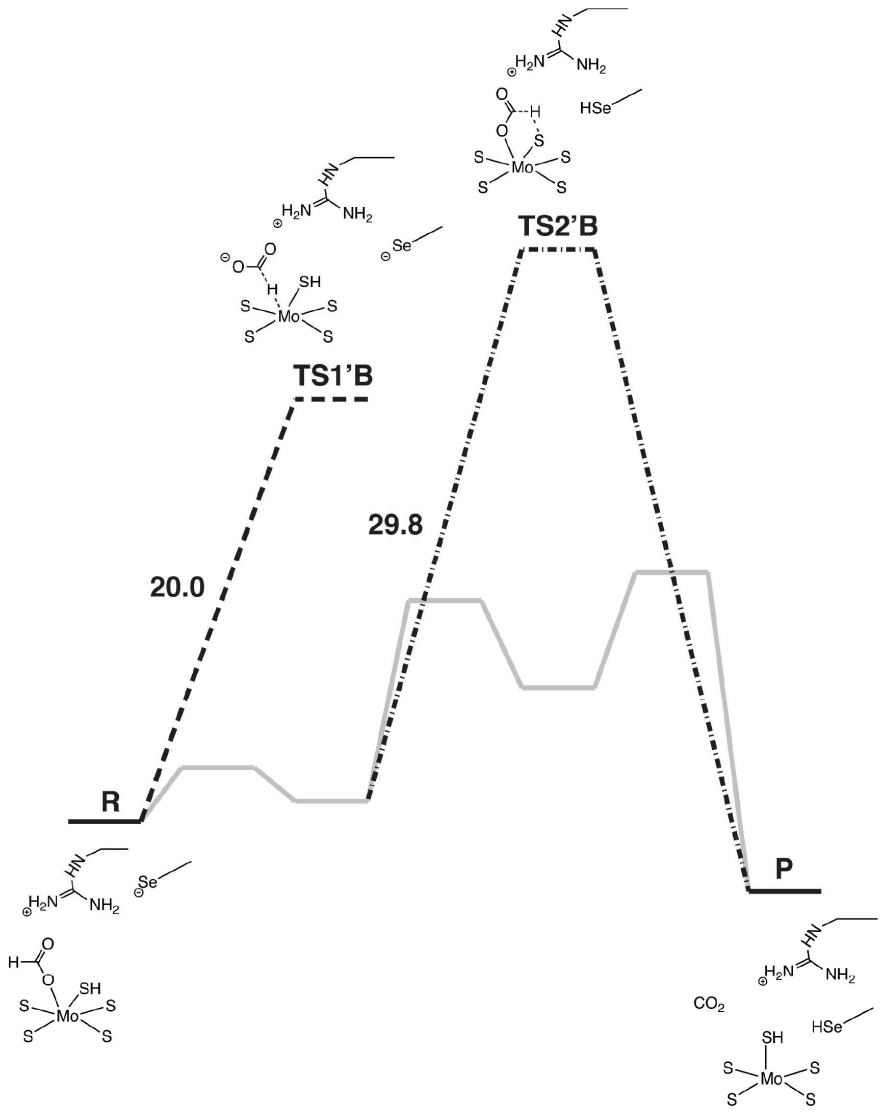

Fig. 3.4 Alternative reaction mechanisms for proton abstraction in model B, Mo-H mechanism. The hydride transfer to Mo from the R state (TS1'B; dashed line), features a higher activation barrier of 20.0 kcal mol^{-1} respect to the same transfer performed from the R'B state. The direct proton transfer from formate to S^{2-} has also been tested, and features a relatively high barrier of 29.8 kcal mol^{-1}(TS2'B; dashed and dotted line). The energy profile of the most kinetically favored reaction mechanism, described in the main text, is shown in grey.

reactants (R'B). After TS1B, a metal hydride intermediate IB is formed, in which the S^{2-} ligand is still unprotonated. It is worth noting that two routes alternative to the pathway shown in Figure 3.4 might be followed: first, H transfer from formate in R to Mo in IB might occur without the proton shift from $SH^-$to Se-Cys (see Figure 3.4; TS1'B) and, second, IB formation might be bypassed through a direct HCO_2^--to-S^{2-} proton shift (see Figure 3.4; TS2'B). However, both alternative mechanisms are energetically disfavored relatively to the pathway going through TS1B.

The final H transfer (from Mo to S^{2-}) occurs through TS2B (see Figure 3.7 for structural details), with an activation barrier of 8.2 kcal mol^{-1}; essentially, TS2B is characterized by a slight stretching of the Mo-H and S-H bonds and features a smaller S-Mo-H angle relative to IB. The products formed after TS2B are the same obtained in the SeH pathway. The MoH-B route evidences the role of Mo as a favorable mediator for H transfer. In fact, such a pathway is the most viable among those investigated, featuring a global energy barrier of 14.1 kcal mol^{-1} (corresponding to the maximum energy span associated with the reaction profile; see Figure 3.8, E_{TS2B} - E_R). B3LYP calculations have been subsequently performed to confirm the obtained results also by employing a different (hybrid) functional for the treatment of the electron exchange correlation. The resulting energy profiles (see Figures 3.5 and 3.6) are in full agreement with the general outcome, which shows that metal-mediated pathways for H transfer are favored over other mechanisms.

NBO/NPA calculations have been performed for evaluating atomic partial charges and their variations upon reaction. In most mechanisms, the Mo charge in the products decreases significantly (by 0.30-0.40 e) compared to that observed in the reactants, consistent with reduction of the Mo^{VI} ion. The total charge of formate increases by 0.44-0.60 e on going from the coordinated HCO_2^- to CO_2, implying that the electron density lost by formate during its oxidation is not completely recovered by the Mo ion. In fact, NPA shows the electron density coming from HCO_2^-oxidation is also spread on the pyrazine moiety of the cofactor, thus remarking the importance of a large-sized representation of the MGDs in DFT calculations [138, 177]. These findings fit well also with the

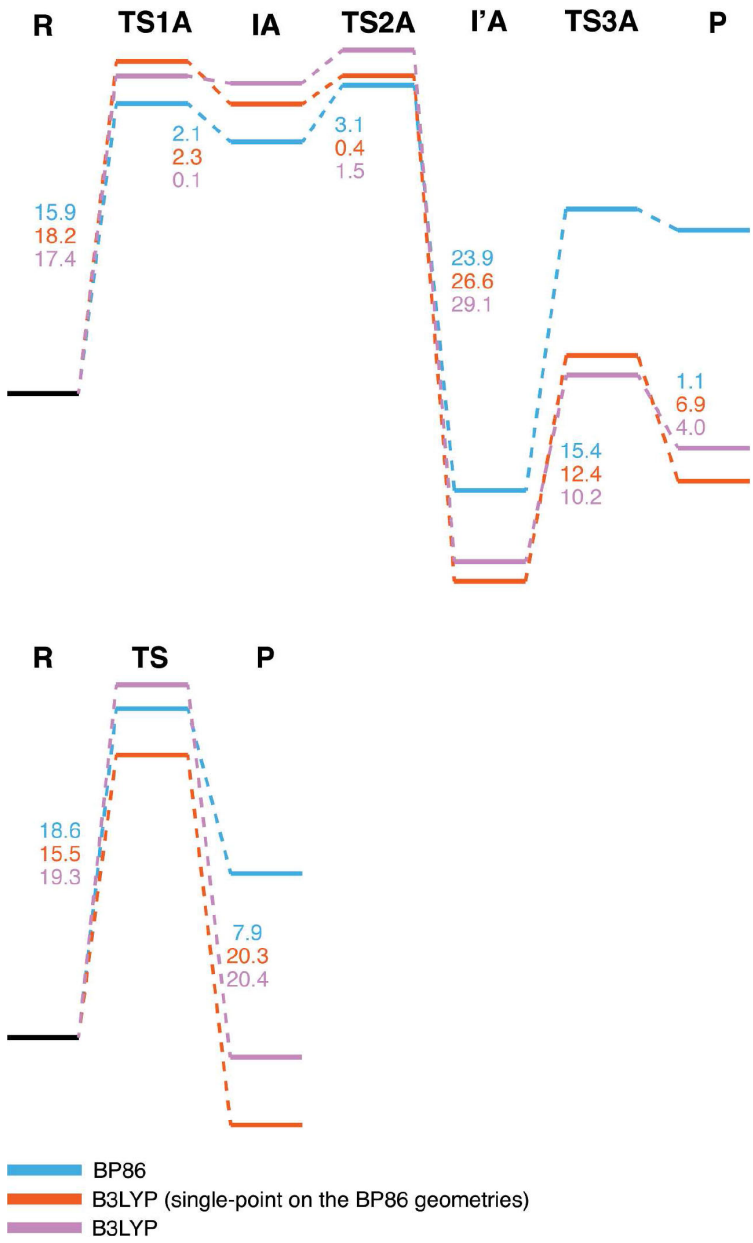

Fig. 3.5 Comparison between MoH A (top) and SeH A (bottom) mechanisms, as resulting from different exchange- correlation functionals employed. B3LYP and BP86 are in full agreement in showing that the H transfer pathway occurring through the metal ion mediation is slightly favored over the other pathway (occurring as a single proton transfer event directly from formate to Se-Cys)

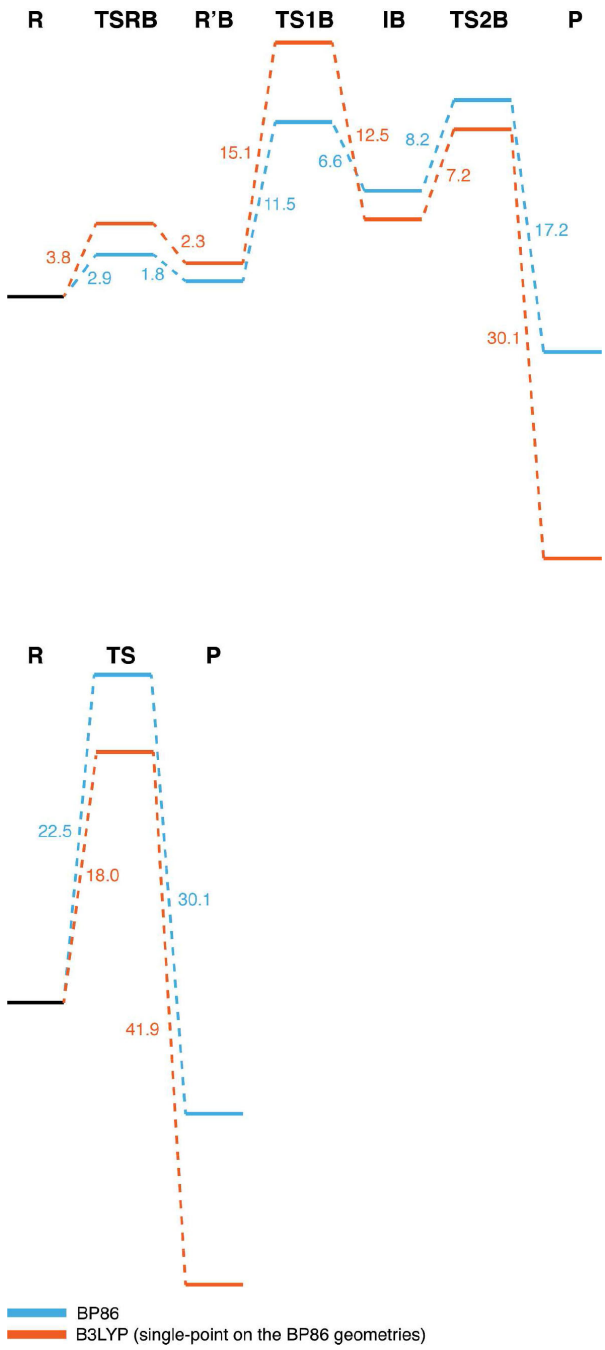

Fig. 3.6 Comparison between MoH B (top) and SeH B (bottom) mechanisms, as resulting from different exchange- correlation functionals employed. B3LYP (single point SCF on the BP86 geometries) and BP86 are in full agreement in showing that the H transfer pathway occurring through the metal ion mediation is favored over the other pathway (occurring as a single proton transfer event directly from formate to Se-Cys)

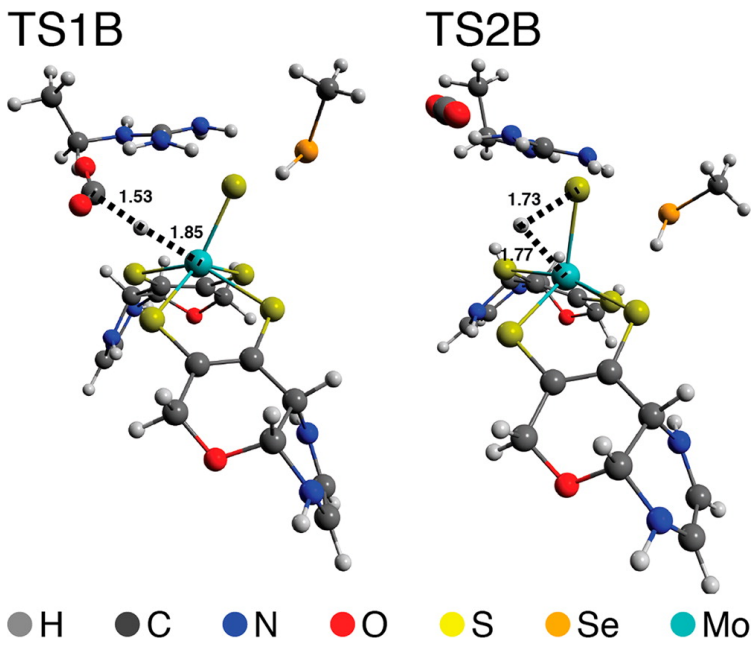

Fig. 3.7 Relevant transition states of the mechanism MoH-B. Structural elements forming the reaction coordinate are in angstroms.

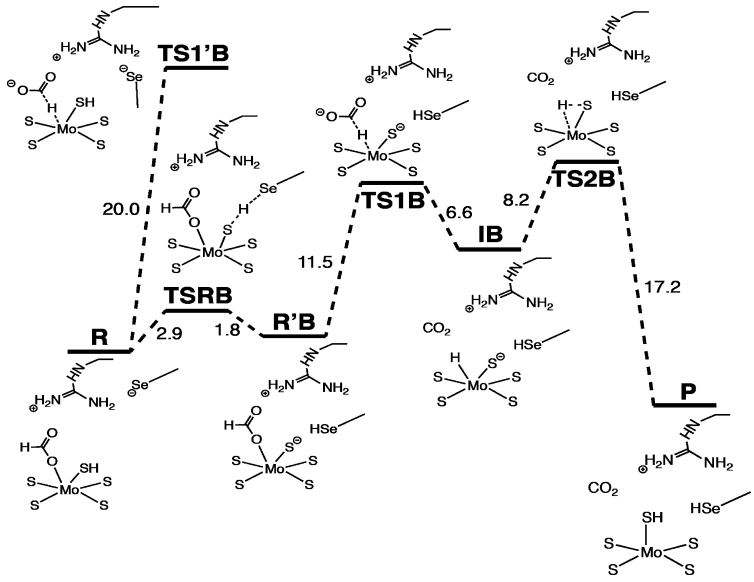

Fig. 3.8 Energy profile (kcal mol^{-1}) of the MoH-B reaction pathway

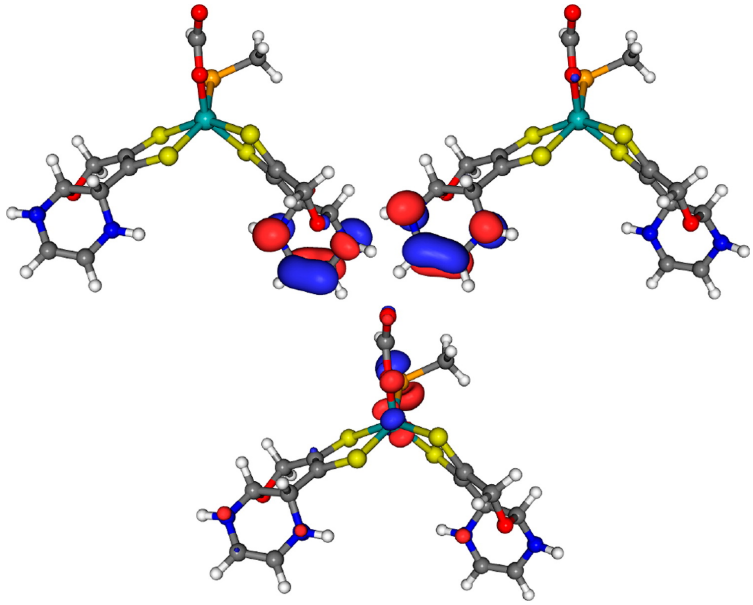

Fig. 3.9 Comparison between MoH B (top) and SeH B (bottom) mechanisms, as resulting from different exchange- correlation functionals employed. B3LYP (single point SCF on the BP86 geometries) and BP86 are in full agreement in showing that the H transfer pathway occurring through the metal ion mediation is favored over the other pathway (occurring as a single proton transfer event directly from formate to Se-Cys)

electron transmission mechanism proposed for FDH-H because the MGD conjugated ring system represents the minimum path for the electrons to reach the Fe4S4 cluster, which is found in the FDH-H X-ray crystal structure [32]. Frontier orbital analysis (Figure 3.9) performed on the cofactor model shows that, in most cases, the LUMO involves only the Mo ion and some coordinating atoms. The HOMO (SOMO for triplet states) is mainly located on the rings of the pyranopyrazine moiety (Figure 3.9), confirming once again the importance of a reliable modeling of the MGD ligand.

3.5 Conclusions

In conclusion, DFT has been used to investigate systematically all possible intimate mechanisms associated with the rate-determining step of the $HCO_2^- \rightleftharpoons CO_2 + H^+ + 2\,e^-$ reaction, catalyzed by Mo-binding FDH. The sampling completeness of the mechanistic investigation focused on the H-transfer step is guaranteed by the limited number of possible H acceptors within the FDH active site. Actually, the acceptor range is composed by the S ligands of MGD, the selenide of the Se-Cys, and the Mo ion. An accurate model of the MGD ligand has been employed, in order to fulfill recently emerged requirements for the correct treatment of the cofactor stereoelectronic features [138, 177]. Among all investigated mechanisms, those featuring an initial hydrogen migration as hydride from the substrate to the metal ion have turned out to be the lowest in energy, regardless of the model employed (MA vs MB, with the latter featuring the absolute lowest-energy barriers) for the enzyme active site. Notably, the conserved Se-Cys retains an important role because it is involved in shuttling protons from and to the metal cofactor. The proposal of the possible role played by the Mo ion as a mediator of H transfer from (to) formate to (from) the Se-Cys has been reinforced by using two different among the most popular (and generally reliable) electron density functionals, namely, the pure BP86 and the hybrid B3LYP. The observation that part of the electron density arising from formate oxidation is localized on the pyrazine moiety of the cofactor, as well as localization of HOMO/SOMO on the same region, confirms that employing an extended model of the cofactor is mandatory for describing properly the chemistry of FDH and MPT binding enzymes in general.

Examples of high-valent molybdenum hydride complexes have already been reported [7, 39, 44]. In particular, even though rare, thiolate-coordinated molybdenum hydrides have been spectroscopically characterized [52, 53]. Moreover, group 6B carbonyl hydrides, such as $[HM(CO)_5]^-$ (M = Cr, Mo, W), can perform $HCO_2^- \rightleftharpoons CO_2 + H^+ + 2\,e^-$ interconversion [52, 53]. In light of our results, the combination of a low-valent M center with strong acceptor ligands, as observed in $[M(CO)_5]$ complexes,

could be tentatively conceived as analogous to high-valent Mo ligated by electron donors, as in the FDH active site. Actually, the computed Mo partial charges in FDH models and the $[HMo(CO)_5]^-$ complex are comparable. This observation further supports the idea that hydride transfer may be a viable pathway in the FDH catalytic mechanism. Its strong resemblance with a key step of the catalytic mechanism of a long-standing inorganic process, such as the water-gas shift reaction, is noteworthy. Concerning the CO_2 reductive process, the proposed mechanism fulfills the paradigmatic requirement of generating an activated metal hydride, which is able to perform a nucleophilic attack to the C atom [27].

Chapter 4

Hydrogen oxidation in [NiFe]-hydrogenase

4.1 Introduction

H_2 plays a central role in many chemical processes, and it has promising potential as an environmentally friendly energy carrier. However, a key requirement for the realization of large-scale and sustainable processes, in which H_2 is used for energy storage and delivery, is the development of new catalysts based on inexpensive metals [213].

Hydrogenases are enzymes involved in the metabolism of H_2 and can be classified, according to the composition of their active site, into three families: [NiFe]-, [FeFe]-, and [Fe]-hydrogenases [160, 205]. [NiFe]- and [FeFe]-hydrogenases are classical hydrogenases and catalyze the reversible oxidation of H_2 [2, 102]. Since in vivo [NiFe]-hydrogenases are generally involved in H_2 oxidation whereas most [FeFe]-hydrogenases are involved in H_2 production, the former are very promising targets for reverse engineering studies aimed at disclosing the stereoelectronic features needed for efficient H_2 binding and cleavage, and therefore drive the design of bioinspired catalysts [87, 140, 197].

In the active site of [NiFe]-hydrogenases, the Ni atom is coordinated by four cysteine residues: two in a terminal fashion and two bridged between the Ni and Fe atoms. The Fe atom is also coordinated by three

non-proteic ligands, identified by FTIR and X-ray studies as two CN^- and one CO. A third ligand bridges the two metal atoms in several enzyme forms: an oxygen-containing ligand in oxidized inactive forms and a hydride in reduced active forms [76, 95, 103, 154, 164, 208–210]. Therefore, the coordination environment of Fe and Ni in these enzyme forms can be described as octahedral and distorted trigonal bipyramidal, respectively. The characterization of several forms of [NiFe]-hydrogenases, differing for their redox and/or protonation state, allowed also to outline key features of the enzyme reactivity (Figure 4.1) [151, 154, 188].

The oxidized states of [NiFe]-hydrogenases are catalytically inactive and generally referred to as Ni-A and Ni-B. One-electron reduction of the Ni-A and Ni-B forms yields EPR-silent species referred to as Ni-SU and Ni-SI_r, respectively. Protonation of Ni-SI_r and subsequent dissociation of the oxygen containing ligand from the active site yields the Ni-SI_a, form, which has been suggested to correspond to the redox species that binds and activates H_2. In Ni-SI_a, both the Ni and Fe atoms have the formal $Ni^{II}Fe^{II}$ oxidation state, and are four- and five-coordinated, respectively; i.e., the metal ions are both characterized by one vacant coordination position, where H_2 might bind. Reduction and protonation of the Ni-SI_a form yields a catalytically active paramagnetic form, usually referred to as Ni-C, which corresponds to a $Ni^{III}Fe^{II}$ species characterized by a hydride ligand bridging the two metal ions. One electron reduction of Ni-C leads to the most reduced active form of the enzyme, which is usually referred to as Ni-R. Experimental and computational results converge in the description of the Ni-R form as a species featuring a hydride ligand simultaneously coordinated to Ni^{II} and FeII. In addition, three distinct forms of Ni-R (Ni-R1, Ni-R2, Ni-R3), which differ for the protonation state, have been identified by IR spectroscopy and theoretical calculations [54, 55, 71].

Even if several [NiFe]-hydrogenase redox forms have been characterized (Figure 4.1), some key details about the reactivity of the enzyme are still fuzzy. One ambiguity is related to the spin state and coordination geometry of Ni^{II} in the Ni-SI_a state of the enzyme. In general, when considering sterically unstrained species, four-coordinated low spin Ni^{II} (S =

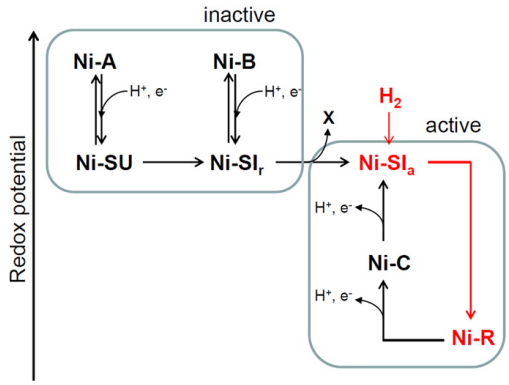

Fig. 4.1 Activation Mechanism and Catalytic Cycle of [NiFe]-Hydrogenase. X is an oxygen containing species (probably H2O) which leaves the active site in the activation process. The part of the catalytic cycle explicitly investigated in this work is shown in red.

0) complexes are expected to be characterized by square planar geometry, whereas high spin Ni^{II} (S = 1) species should have tetrahedral geometry. In fact, in most crystallographic structures of [NiFe]-hydrogenases, the Ni atom is five-coordinated, due to the presence of a non-amino-acidic bridging ligand (referred to as X in figure 4.2), and features distorted trigonal bipyramidal geometry. Removal of the exogenous bridging ligand, as occurs when the Ni-SI_a state of the enzyme is formed, might leave the Ni^{II} atom in a peculiar seesaw coordination geometry, if the protein backbone does not allow relaxation of the Ni ligands (figure 4.2).

Another ambiguity is related to the site of H_2 binding. In fact, a species in which H_2 is coordinated to the enzyme active site, but not yet cleaved, is expected to be very short-lived and has not yet been characterized experimentally. Initial DFT calculations carried out using simple models of the $Ni^{II}Fe^{II}$ Ni-SI_a form of the active site showed that H_2 can be weakly bound to the five-coordinated Fe^{II} atom, whereas the corresponding adduct between the Ni^{II} ion and H_2 is unstable [148]. However, this observation was difficult to reconcile with the experimental observation that the gas channel connecting the enzyme surface with the active site ends in front of the Ni atom [143]. Notably, more recent DFT results, obtained using computational models of the active site in which

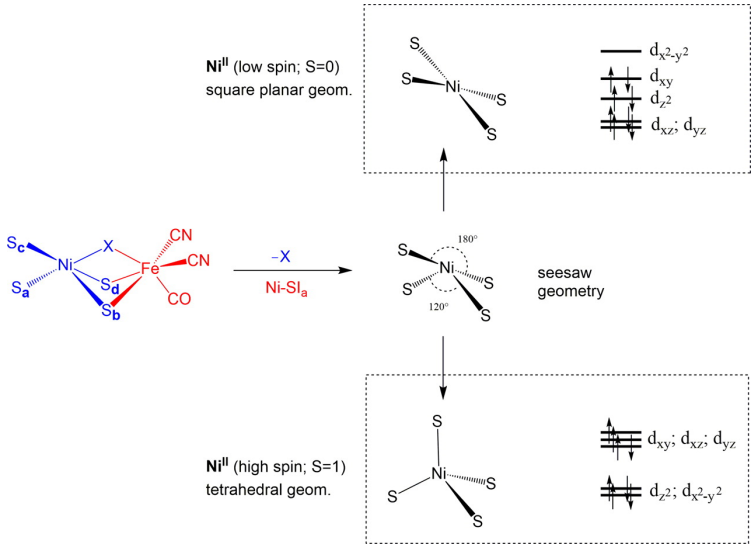

Fig. 4.2 Coordination Geometry of Ni^{II} in the NiFe Cluster. Removal of the bridging ligand X might leave the Ni^{II} ion in a seesaw geometry, which differs from the classical square planar or tetrahedral geometries expected for low or high spin Ni^{II} complexes, respectively.

the positions of the four cysteine residues forming the Ni coordination environment were constrained to the corresponding positions observed in the X-ray structure of the enzyme, indicated that both Ni–H_2 and Fe–H_2 adducts correspond to plausible intermediate species in the catalytic mechanism [220].

According to experimental and computational results, H_2 cleavage should eventually lead to the formation of a species in which one hydrogen atom ends up as a bridging hydride and the other as a proton bound to a terminal cysteine. However, the redox state of the Ni ion when H_2 cleavage takes place is still controversial. In fact, also in this case, computational results depend to some extent on the chemical models adopted in the calculations: H_2 cleavage is predicted to be more likely on a $Ni^{III}Fe^{II}$ species when small models of the NiFe cofactor are considered, whereas calculations carried out using larger models (i.e., including the effect of the environment of the NiFe cofactor) suggest that H_2 cleavage can already take place at the $Ni^{II}Fe^{II}$ redox state [131, 148, 188, 193, 220].

Previous computational studies explored also a scenario where H_2 could bind and be cleaved on an active site form in which a hydride ligand is already bridged between the two metal ions [8]. As far as Ni^{II} is assumed for H_2 cleavage, parallel mode EPR [59], saturation magnetization [211], and UV–visible MCD [123] investigations suggest a low spin Ni^{II}, whereas L-edge X-ray absorption spectroscopy [212] suggests a high spin Ni^{II}. In this respect, calculation of relative stabilities by DFT can be of little aid due to its known inherent limitations in predicting the correct ground state among different spin states [36, 173].

4.2 Aim

With the aim of contributing to shed more light on the reactivity of [NiFe]-hydrogenase and, more specifically, disclose which stereoelectronic properties of their active site are crucial for efficient H_2 binding and cleavage, we have adopted a comparative approach, using DFT to study a series of systems ranging from minimal models of the active site, to very large models in which all the aminoacids forming the first and second coordination sphere of the NiFe cluster have been explicitly included.

4.3 Methods

Quantum mechanics (QM) calculations have been carried out in the DFT framework with the TURBOMOLE 6.4 suite of programs [4] by using the BP86 functional [23, 162] in conjunction with the resolution-of-the-identity (RI) technique [64]. In the large- and medium-size models, an all-electron valence triple-ζ basis set with polarization functions TZVP [181] has been used for Ni, Fe, the atoms of the first coordination sphere, and the residue Glu14. For all other atoms, the double-ζ basis set SV(P) [181] has been used. In the minimal model, the TZVP basis set is used for all atoms.

The effects of the protein environment surrounding the models have been modeled according to the conductor-like screening model (COSMO) [119, 120] by considering a polarizable continuum medium with $\varepsilon = 4$. In

the case of the large model, geometry optimizations have been carried out in the polarizable continuum medium, whereas for the medium-size model only single point calculations in the polarizable continuum medium have been performed on the geometries optimized in a vacuum. It is interesting to note that energy differences calculated in a vacuum and in the polarizable continuum medium are very similar, indicating that the size of the models is large enough to be insensitive to nonspecific polarizing effects of the protein environment. The work mainly focuses on H_2 binding and cleavage on the putative $Ni^{II}Fe^{II}$ Ni-SI_a form of the enzyme. Both singlet (S = 0) and triplet (S = 1) spin states for the Ni atom have been considered. For comparative purposes with previous proposed mechanisms, H_2 binding and cleavage have also been investigated on the $NiII^{I}Fe^{II}$ (S = 1/2) and low spin $Ni^{III}Fe^{II}$ (S = 1/2) redox states of the active site.

Geometry optimizations of the species along the $Ni^{II}Fe^{II}$ pathways have been performed by including the dispersion interactions correction given by the DFT-D3 Grimme scheme [91–93] on the large model, as implemented in TURBOMOLE [5], and in order to verify the consistency of results, we have also carried out calculations for the energetically most favored pathways using the M06L functional and the TZVP-SV(P) basis set, on the medium-size model [224]. Calculations with the M06L functional have been performed using the Gaussian 09 program [81]. The results obtained by both including the dispersion interactions correction and using the M06L functional are consistent with those obtained by using the BP86-COSMO scheme (data not shown).

Vibrational analysis has been carried out for all species optimized using the medium-size model. Due to the atoms constrained at the original positions, several very small negative eigenvalues of the Hessian were computed. These eigenvalues correspond to torsional vibrational modes involving the constrained atoms.

Transition state search has been performed according to a pseudo Newton–Raphson procedure. First, a geometry optimization of a transition state guess geometry, in which selected internal coordinates are constrained, was performed. The Hessian of such an optimized geometry was then computed, and if one negative eigenvector corresponding to

the reaction coordinate was found, the geometry was used as the starting point for the successive optimization steps applying the trust-region image minimization method [97], which is designed to maximize the energy along one of the eigenvectors (i.e., the chemically relevant one), while minimizing it in all other directions. As discussed above, the presence of constrained atoms resulted in the occurrence of several imaginary frequencies. However, the eigenvector corresponding to the reaction coordinate was easily identified as the one with an eigenvalue much lower than the other negative eigenvalues.

4.4 Results and discussion

Throughout this chapter, computational models will be labeled according to the general scheme ^{i}K-X, where i = spin state of the complex, K = formal oxidation state of the Ni ion, and X = specific chemical nature of the complex. For the sake of clarity, when necessary, the size of the model (S, M, and L for small, medium, and large models, respectively) will be indicated as a subscript after K.

The starting structure for the DFT calculations was based on the X-ray geometry of the *Allochromatium vinosum* (PDB ID: 3MYR) [NiFe]-hydrogenase [150], in which the oxo ligand bridging the Ni and Fe ions was removed, in order to obtain the Ni-SI$_a$ coordination environment, where Ni and Fe are four- and five-coordinated, respectively. In the following, the residues are numbered according to the 3MYR structure (all residues included in the models belong to the large chain of the enzyme, with the exception of Thr17 which belongs to the small chain and, for the sake of clarity, is labeled as Thr17(S)). Three models of different size have been considered to investigate the effect of the protein environment on the stereoelectronic properties of the active site of [NiFe]-hydrogenases. The validity of such an approach (cluster models) for the DFT investigation of transition metal containing enzymes has been extensively discussed [6, 186, 187].

The largest model, which is schematically shown in figure 4.3, contains up to 290 atoms and has a size of about 18 Å. This model includes in its

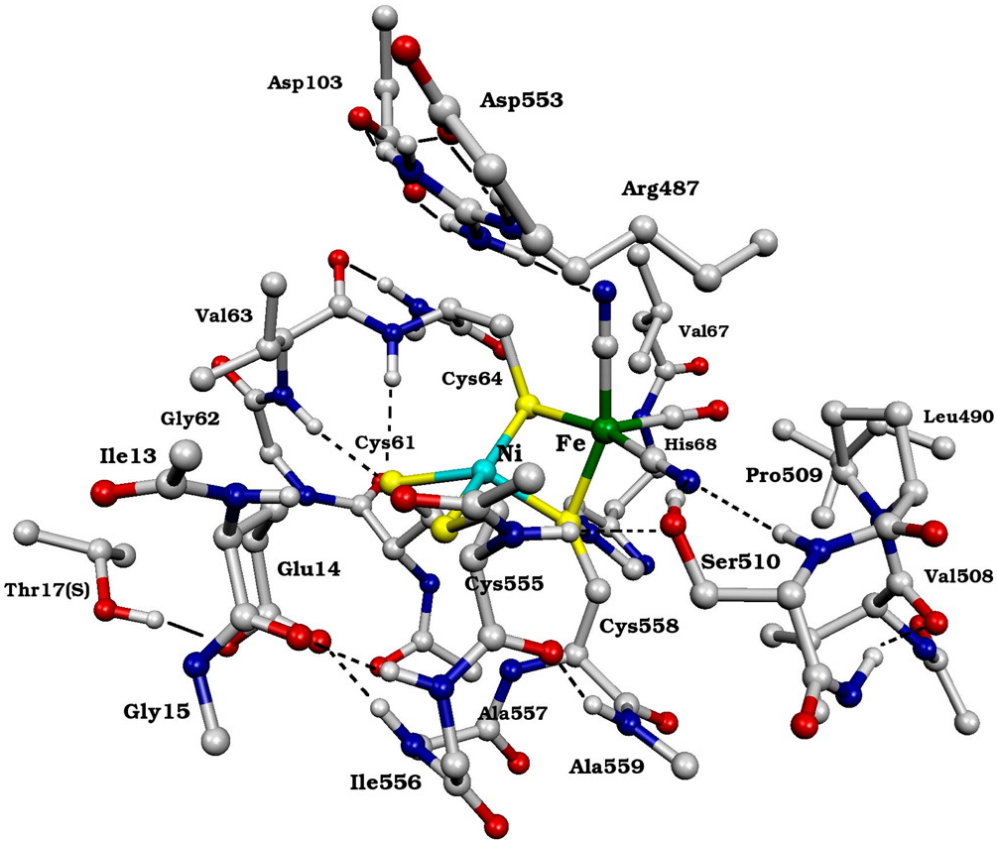

Fig. 4.3 Schematic representation of the largest DFT model of the active site used in this work, derived from the X-ray crystal structure of the [NiFe]-hydrogenase from *Allocromatium vinosum* (pdb code 3MYR)

core the bimetallic NiFe cluster, with two CN^- and one CO ligands coordinated to the Fe atom, two cysteine residues terminally coordinated to Ni, and two further cysteine residues bridging the two metal atoms. In the enzyme, two pairs of terminal and bridging cysteine residues (Cys61/Cys64 and Cys555/558) belong to two different Ni binding motifs located at the N- and C-termini of the large subunit. In each binding motif, the cysteine residues are separated by two other residues, which are also contained in the model, so that the pairs Cys61/Cys64 and Cys555/558 are covalently linked together, forming two peptide chains composed by four residues. For the peptide chain Cys555-Ile556-Ala557-Cys558, only the backbone atoms of Ile556 and Ala557 are included in the model, since their side chains are oriented far from the active site, whereas for the peptide chain Cys61-Gly62-Val63-Cys64 the side chain of Val63 is included in the model, since it is close to the vacant coordination site on Ni. These two peptide chains are terminated by selected backbone atoms of residues Ile60, Thr65, Pro554, and Ala559. The second coordination sphere of the NiFe cluster includes the side chains of residues Thr17(S), Asp103, Leu490, Arg487, and Asp553, terminated at the Cα atoms; a peptide chain containing the residues His68 and Val67; a peptide chain containing the carbonyl and the Cα atoms of Val507, and the entire residues Val508, Pro509, and Ser510; and finally a peptide chain containing the carbonyl and Cα atoms of Ile13, the entire residue Glu14, and the N and Cα atoms of Gly15. The truncated residues were saturated with hydrogens. During geometry optimizations, selected atoms have been constrained to the crystallographic positions, in order to avoid unrealistic distortions at the boundary of the model.

Arg487, Asp553, and Asp103 have been included in the model, since they form a network of H-bonds which also involves one CN^- ligand, whereas Ser510 is H-bonded to the other CN^- ligand through its hydroxyl and amide hydrogens. Val67, Leu490, Val508, and Pro509 have been selected because they form a hydrophobic pocket surrounding the $Fe(CN)_2(CO)$ group.

His68 is close to the bimetallic cluster, and when protonated at the Nε atom, it can form a H-bond with the sulfur atom of the bridging

Cys558. Three protonation states are possible for this residue depending on whether Nδ, Nε, or both atoms are protonated. These protonation states have been investigated in detail by Neese et al. [115], using a model based on the X-ray structure of the *Desulfovibrio vulgaris* Miyazaki F [NiFe]-hydrogenase; it was concluded that the best comparison between experimental and theoretical data is obtained when His68 is protonated at the Nε atom. On the basis of these results, His68 was always modeled as protonated at the Nε atom.

The carboxyl group of Glu14 is close to the sulfur atom of Cys555, and it was proposed to be involved in proton transfer from/to the bimetallic cluster [57, 84, 198]. In the X-ray structure of the enzyme from *Allochromatium vinosum*, it forms H-bonds with Ala557 and Thr17(S), which were also retained in the model. When investigating H_2 binding to the active site, Glu14 has been modeled both as glutamate and glutamic acid. In the latter case, the proton forms a H-bond with Cys555. Selected backbone atoms of the adjacent residues Ile13 and Gly15 were also added to the model, since they form H-bonds with residues Pro554 and Ile556 belonging to the Cys555/Cys558 binding motif.

A sequence alignment of the large chain of several [NiFe]-hydrogenases shows that residues for which the side chain is included in the model are conserved, emphasizing the possible relevance of these residues in the catalytic mechanism of the enzyme. Only backbone atoms of non-conserved residues are included in the model, generally to terminate peptide chains. Exceptions are Val67 and Ser510. The former is replaced by Thr, Ser, Cys, or Ala residues in other [NiFe]-hydrogenases, whereas the latter can be replaced by a Thr residue. A mutagenesis study carried out on the Cys residue replacing Val67 in the [NiFe]-hydrogenase from Ralstonia eutropha showed that the effect of the amino-acid substitution on the activity of the enzyme is rather small [135, 179].

The medium-size model contains up to 122 atoms. This model has been used for optimizations of both minimum energy structures and transition states along the H_2 cleavage pathways. The model contains the first coordination sphere of the NiFe cluster, with the cysteine residues terminated at Cα atoms. In the second coordination sphere, the model includes

the side chains of Asp103, Asp553, His68, and Thr17(S) terminated at Cβ atoms; the side chain of Glu14 terminated at the Cγ atom; and the side chain of Arg487 terminated at the Cδ atom. The model also contains the side chain of Ser510 and the backbone atoms of Ile556 and Ala557, which are involved in a H-bond network with Glu14. For the medium-size model, initial coordinates were taken from the corresponding optimized structure of the large-size model. During geometry optimizations, terminal atoms were constrained to their initial positions. A minimal model including only the ligands of the first coordination sphere of the two metal atoms, with formula $[(CH_3S)_4Ni^{II}Fe^{II}(CN)_2(CO)]^{2-}$, has also been considered. In this model, the four cysteine residues are terminated at the Cβ atoms and they are therefore modeled as methylthiolates.

Energy profiles as a function of the S_a–Ni–S_b angle (see figure 4.3 for atom labeling) have been calculated for the small model $[(CH_3S)_4Ni^{II}Fe^{II}(CN)_2(CO)]^{2-}$, which includes only the ligands of the first coordination sphere (^12_S). The model labeled results from a constrained geometry optimization of ^12_S, in which all atoms with the exception of hydrogens have been fixed at the position computed for the large model (see the Results and Discussion for model labeling). It should be noted that full geometry optimization of ^12_S does not lead to the absolute minimum, because rotation of the methyl groups leads to a conformation lower in energy by 1.1 kcal mol^{-1}. However, we will consider ^12_S as the reference geometry for the square planar form, since it provides the smallest reorganization of the pendant groups when moving from ^12_S-P to ^12_S. Geometry optimizations of ^12_S have also been carried out constraining the S_a–Ni–S_b angle at selected values from 160 to 100°. The same protocol has been applied to the H_2-bounded form of ^12_S (^12_S-NiH_2). The H_2 binding energy as a function of the S_a–Ni–S_b angle was then calculated as the energy difference between products and reactants for the reaction $^12_S + H_2 \longrightarrow {}^12_S$-$NiH_2$in which ^12_S and ^12_S-NiH_2 feature the same S_a–Ni–S_b angle. Energy profiles as a function of the Sa–Ni–Sb angle have also been calculated on the mononuclear model $[(CH_3S)_4Ni^{II}]^{2-}$, which includes the Ni ion and its first coordination sphere, using the same procedure described above.

Initially, with the aim of characterizing the H_2 binding site in [NiFe]-hydrogenases, we have studied the interaction of H_2 with the Ni-SI_a redox form of the enzyme. Both low and high spin states have been taken into account. The computational model used in this investigation has been derived from the crystallographic structure of the [NiFe]-hydrogenase from *Allochromatium vinosum*, which should correspond to the Ni-B form, and therefore features an oxo-containing ligand which bridges the two metal atoms in the active site. The removal of this ligand from the crystallographic structure, to generate models of the Ni-SI_a form, leaves the Ni atom in a distorted seesaw conformation in which the Sa–Ni–Sb and Sc–Ni–Sd angles are 109.2 and 165.8°, respectively (see figure 4.2 for atom labeling).

Geometry optimization of a large model of the Ni-SI_a form (see the Methods for details), either in the low or high spin state (^12_L and ^32_L, respectively), does not lead to a significant structural rearrangement of the Ni coordination sphere relative to the X-ray structure. In particular, the Sa–Ni–Sb and Sc–Ni–Sd angles in the geometry optimized models are 123.9 and 150.7°, respectively, for the low spin Ni^{II} model and 121.9 and 155.5°, respectively, for the high spin Ni^{II} model. In order to check the influence of the constraints imposed to the terminal atoms of the model on the Ni coordination geometry, we also performed geometry optimization after removing all of the constraints on the atoms which are covalently connected to the metal cluster (i.e., all constraints on atoms of the two Ni-binding motifs). However, the geometry of the Ni atom remained essentially identical to the one obtained from the corresponding constrained geometry optimization.

The role of the protein environment in tuning the coordination geometry of Ni in the Ni-SI_a form is highlighted by the analysis of the geometries obtained using a medium-size model (see the Methods for details). In particular, for the low spin species (^12_M), the Sa–Ni–Sb angle is significantly larger than in the corresponding large model (S_a–Ni–S_b = 137.8°, Sc–Ni–Sd = 160.4°), indicating that the conformation is approaching the square planar geometry, whereas for the high spin species (^32_M) the Sc–Ni–Sd angle decreases considerably (S_a–Ni–S_b = 110.0°, S_c–Ni–S_d

= 106.0°), indicating that the conformation is approaching the tetrahedral geometry.

In light of the results obtained with large- and medium-sized models, it can be suggested that the seesaw conformation of the bimetallic cluster in the Ni-SI_a form may be favored by the network of H-bonds occurring nearby the active site. In particular, the Cys555/Cys558 binding motif is included in a small α-helix at the C-terminus of the large subunit, which may be crucial for the coordination geometry of the thiol ligands. In addition, the H-bonds between His68/Cys558, Glu14/Pro554, Ile556/Glu14, and Cys555/Ser510 can impart further rigidity to the structure of the active site.

The computed energy difference between the low and high spin forms of Ni-SI_a, calculated using the large model, is 6.7 kcal mol^{-1}in favor of the former. However, due to the known inherent limitations of DFT in predicting the correct ground state among different spin states [36, 173], we cannot consider this value as reliable for the correct assignment of the spin state of the Ni-SI_a form. More relevantly for the aim of our study, binding of H_2 to the low spin Ni^{II} atom in $^1\mathbf{2}_L$ is favorable (-4.0 kcal mol^{-1}; $^1\mathbf{2}_L$-NiH_2; see figure 4.4), whereas the corresponding adduct in which H_2 binds to the Fe^{II} ion is unstable, resulting in H_2 release from Fe and binding to Ni during geometry optimization. As for the high spin $Ni^{II}Fe^{II}$Ni-SI_a model, H_2 can bind to both Ni and Fe ions, but in both cases, the reaction is endoenergetic by 5.3 and 3.7 kcal mol^{-1}, respectively ($^3\mathbf{2}_L$-Ni_{H_2} and $^3\mathbf{2}_L$-Fe_{H_2} in figure 4.4). Therefore, the metal atom featuring the highest affinity for H_2 is the Ni ion in the low spin $Ni^{II}Fe^{II}$redox state of the enzyme. $^1\mathbf{2}_L$-Ni_{H_2} is characterized by a distorted trigonal bipyramidal geometry, with the Sa–Ni–Sb and Sc–Ni–Sd angles equal to 110.3 and 174.5°, respectively, and with the H_2 atoms at 1.63 and 1.64 Åfrom the Ni ion.

The investigation of H_2 binding using medium-size models highlights again the noninnocent role of the protein environment. In fact, binding of H_2 to the low spin NiII atom in the medium-size model ($^1\mathbf{2}_M$-Ni_{H_2}) is endoenergetic by 2.4 kcal mol^{-1}. In addition, H_2 does not bind to Ni in the high spin species, whereas it can still bind to Fe ($^3\mathbf{2}_M$-Fe_{H_2}) but with

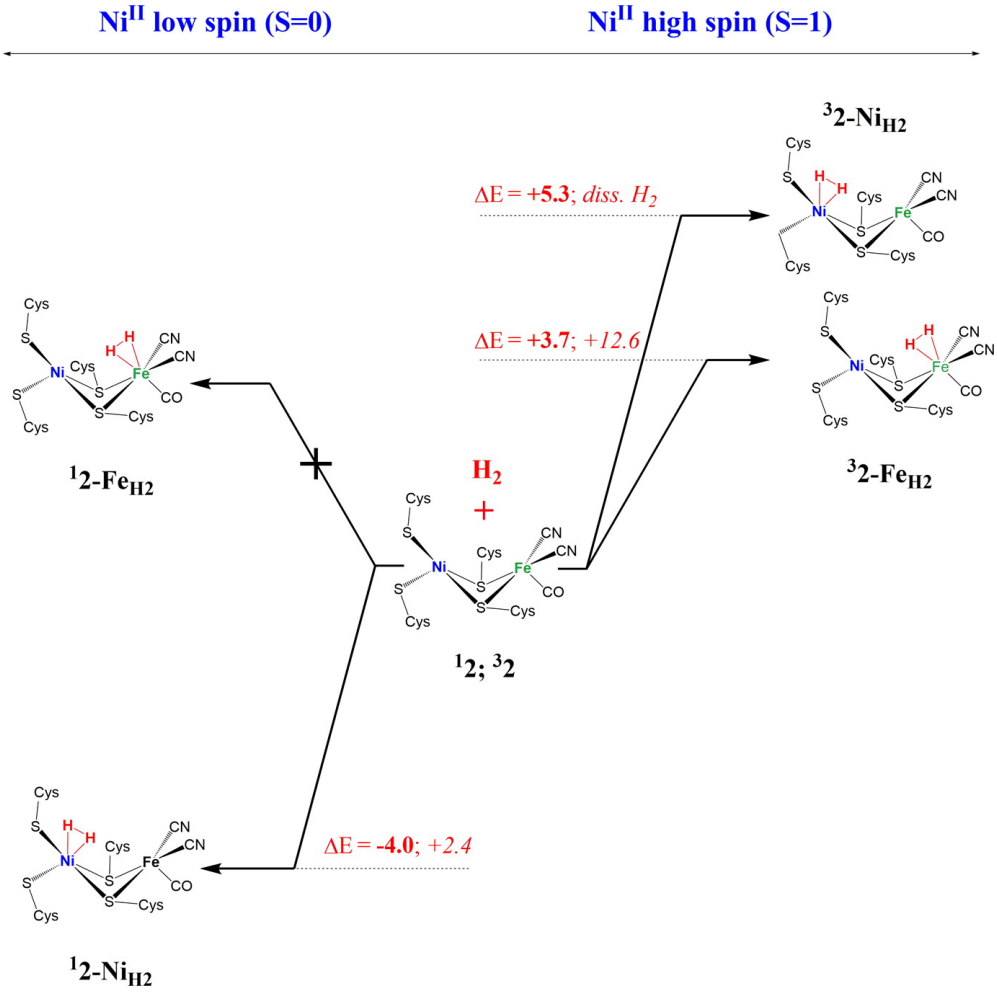

Fig. 4.4 Schematic Representation of the Possible Reaction Pathways for H2 Binding to the Ni-SI$_a$ Form of the Active Site of [NiFe]-Hydrogenases. Values in bold refer to calculations performed using the large model, whereas values in italic refer to calculations performed using the medium-size model. Energies are in kcal mol^{-1}

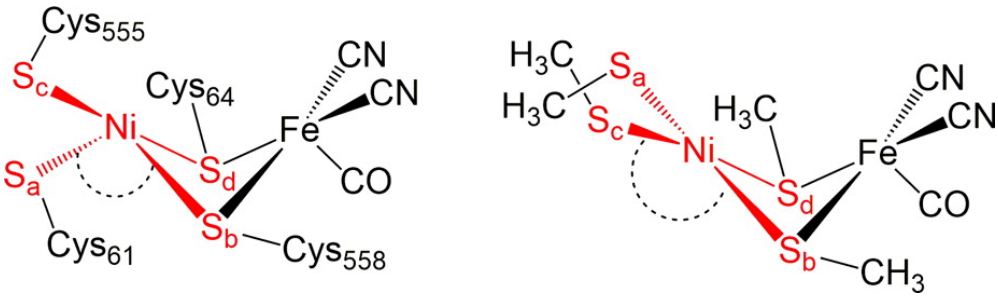

Fig. 4.5 Distorted Seesaw Geometry of Ni in the Large Model ^{1}2$_L$ Corresponding to the Ni-SI$_a$ State of the Enzyme (left) and Distorted Square Planar Geometry in the Small Model ^{1}2$_S$ (right). Smooth transition between the two geometries can be obtained varying the angle S_b–Ni–S_a.

a reaction endoenergetic by 12.6 kcal mol^{-1}.

With the aim of revealing which stereoelectronic factors are crucial for the observed regiochemistry of H_2 binding, we have studied also minimal models of the [NiFe]-hydrogenase active site, which include only the first coordination sphere of the metal ions. The unrestrained optimization of the low spin $[(CH_3S)_4Ni^{II}Fe^{II}(CN)_2(CO)]^{2-}$ model (^{1}2$_S$) yields a structure which differs from the corresponding optimized structure of the large model ^{1}2$_L$ mainly for the S_a–Ni–S_b angle (S_a–Ni–S_b and S_c–Ni–S_d = 160.4 and 168.8°, respectively in ^{1}2$_S$). In other words, the Ni coordination sphere can be described as having a distorted square planar geometry in the small model ^{1}2$_S$, differently from the distorted seesaw geometry observed in ^{1}2$_L$ (see figure 4.5).

H_2 cannot bind to the Ni ion in ^{1}2$_S$ (the geometry optimization of the H_2 adduct leads to dissociation of H_2), whereas it can bind to the Fe atom (ΔE= -1.9 kcal mol^{-1}). On the other hand, in a ^{1}2$_S$ model variant (hereafter labeled as ^{1}2$_S$-P) in which all atoms with the exception of hydrogens were constrained to the positions calculated for the large model ^{1}2$_L$, H_2 can bind the Ni atom (ΔE = -1.2 kcal mol^{-1}), whereas H_2 binding to the Fe atom does not take place. Therefore, the comparison between unconstrained and constrained small models of the [NiFe]-hydrogenase active site clearly highlights that the geometry of the ligands coordinated to the Ni atom is a crucial factor for the regiochemistry of H_2 binding.

To better evaluate the effect of the Ni coordination geometry on the regiochemistry of H_2 binding, we have also studied the energetics of H_2 binding as a function of the Sa–Ni–Sb angle, carrying out a series of geometry optimizations of ^{1}2$_S$ and the corresponding H_2 adduct (^{1}2$_S$-Ni$_{H_2}$), in which the S_a–Ni–S_b angle was kept fixed at several values in the range 100–160°. It turned out that the Ni–H_2 adduct can be formed only when the S_a–Ni–S_b angle is lower than 135°. As shown in figure 4.6, the corresponding reaction energy varies from about +3 kcal mol^{-1}when the S_a–Ni–S_b angle is 135° to about -5 kcal mol^{-1}when the angle is 102.5°. Correspondingly, the energy of ^{1}2$_S$ increases with the bending of the Sa–Ni–Sb angle (see figure 4.6). In particular, ^{1}2$_S$-P is about 11 kcal mol^{-1}higher in energy than the unconstrained ^{1}2$_S$ model. Therefore, the data reported in figure 4.6 highlight a trade-off between the H_2 binding energy, which becomes more favorable for lower angles, and the destabilization of the cofactor, which increases for lower angles. Finally, we have also studied H_2 binding to the minimal $[(CH_3S)_4Ni]^{2-}$ model, which is obtained by removing the Fe(CN)2(CO) moiety from ^{1}2$_S$, to evaluate how the Fe(CN)2(CO) moiety affects H_2 binding. $[(CH_3S)_4Ni]^{2-}$ can bind H_2 only when the S_a–Ni–S_b angle is lower than 130°, and the H_2 binding energy is systematically less favorable by about 3 kcal mol^{-1}than in ^{1}2$_S$. In addition, as shown in figure 4.6, the energy of the $[(CH_3S)_4Ni]^{2-}$ complex as a function of the Sa–Ni–Sb angle increases more steeply than for ^{1}2$_S$.

To rationalize the effect of the Sa–Ni–Sb angle on the regiochemistry of H_2 binding, we have analyzed the frontier molecular orbitals (FMOs) of ^{1}2$_S$ (see figure 4.7 and 4.8).

In the unconstrained ^{1}2$_S$ structure, the LUMO is mainly localized on the Fe atom and features a lobe with an orientation well suited to act as an electron acceptor in H_2 binding (figure 4.7), whereas the HOMO can be described as an antibonding combination of the p orbitals of the terminal S atoms and one d orbital of the Ni atom (see figure 4.8). When considering the FMOs of ^{1}2$_S$-P, the LUMO is mainly localized on the Ni atom and a lobe is oriented in the region between the two metal atoms. Inspection of the orbital energies reveals that only the LUMO energy is

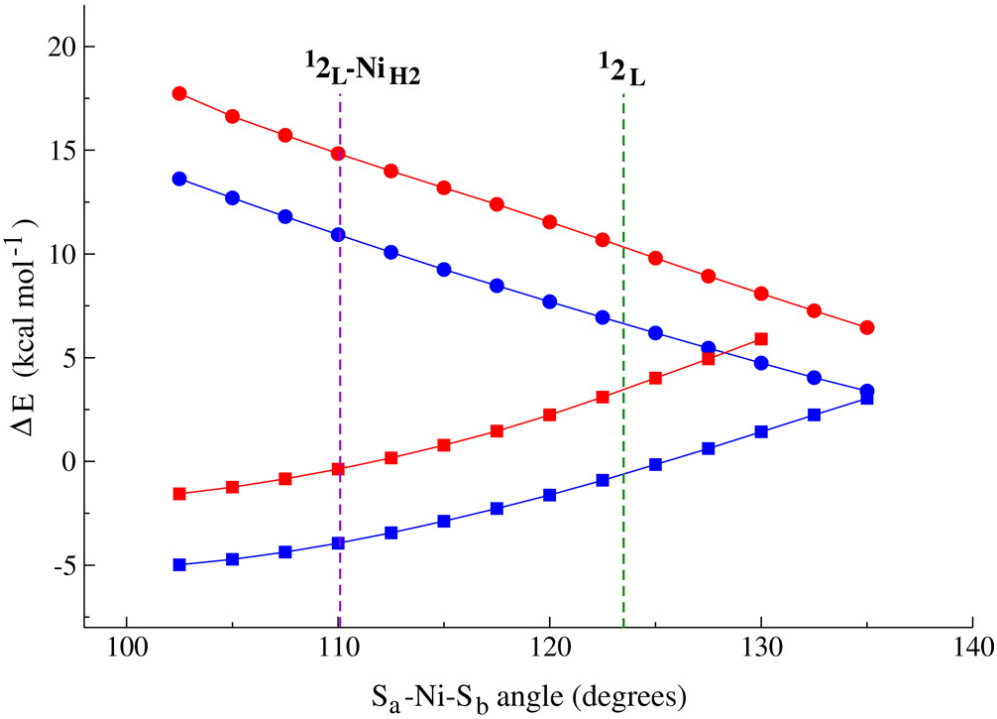

Fig. 4.6 H_2binding energy to Ni calculated as a function of the S_a–Ni–S_b angle for the complex ^{1}2$_S$ (blue line with square symbols) and the complex $[(CH_3S)_4Ni]^{2-}$ (red line with square symbols). Energy variation calculated as a function of the S_a–Ni–S_b angle with respect to the optimum geometry for the complex ^{1}2$_S$ (blue line with circle symbols) and the complex $[(CH_3S)_4Ni]^{2-}$ (red line with circle symbols). Only results in the range 100–135° are reported, because H_2does not bind to Ni for larger values of the angle. Vertical green and violet dashed lines indicate the values of the S_a–Ni–S_b angle measured in ^{1}2$_L$ and ^{1}2$_L$-ligandNiH_2, respectively.

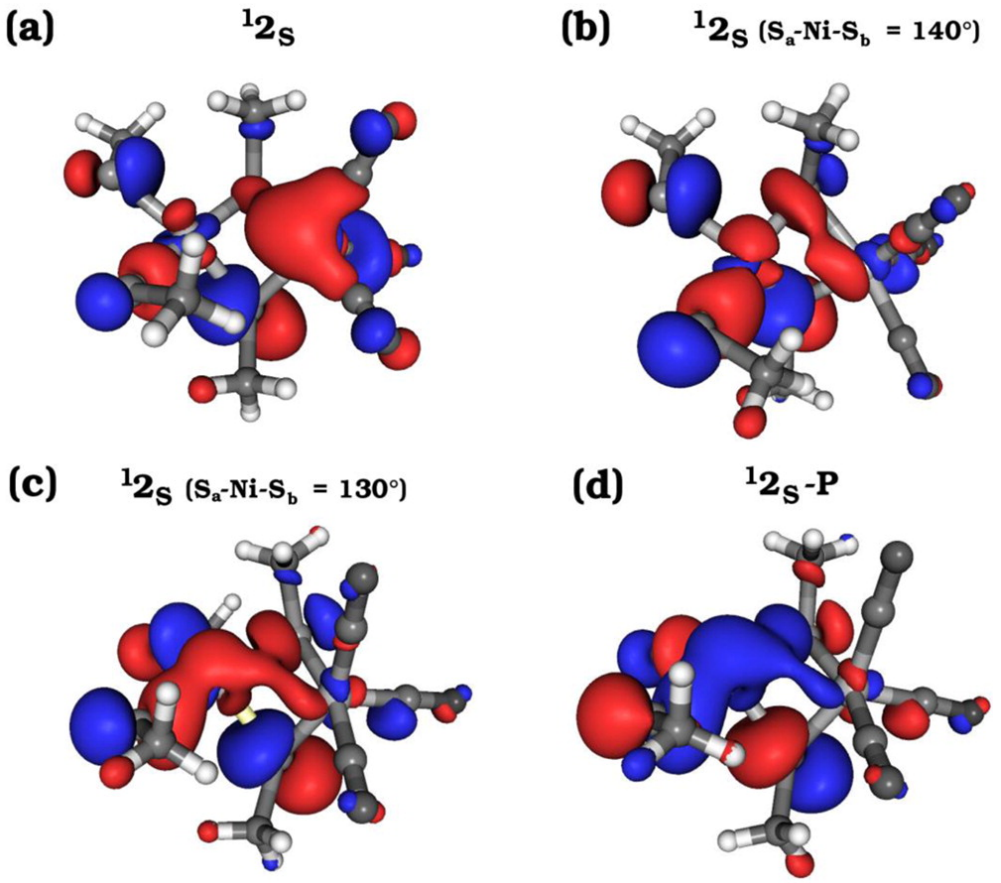

Fig. 4.7 LUMO of ^{1}2$_S$ (a) in the fully unconstrained optimized structure; (b, c) in the structures with the S_a–Ni–S_b angle constrained to 140 and 130°, respectively; and (d) in the model variant ^{1}2$_S$-P$_i$n which the Ni ligands were constrained to the positions observed in the enzyme.

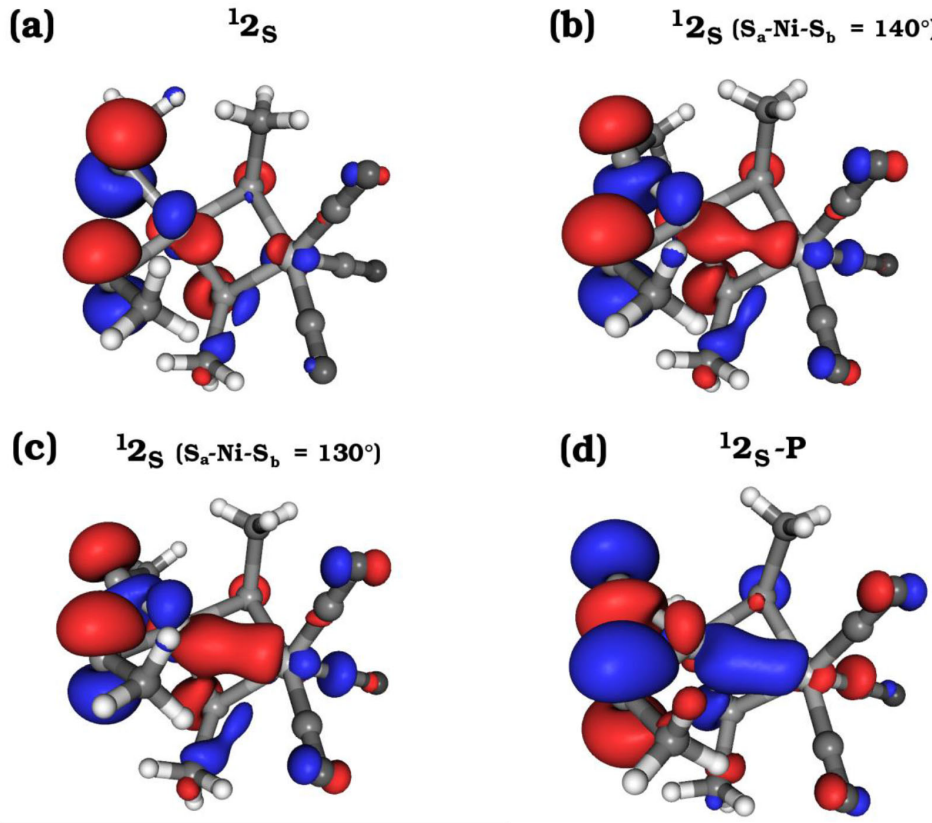

Fig. 4.8 HOMO of ^{1}2$_S$ (a) in the fully unconstrained optimized structure; (b-c) in the structure with the S_a-Ni-S_b angle constrained to 140 and 130 degrees, respectively; and (d) in the model variant 1 ^{2}S–P_in which the Ni ligands were constrained to the positions observed in the enzyme.

affected by variation of the S_a–Ni–S_b angle, decreasing when the bending of this angle increases. The localization of the LUMO on the two different metal atoms in $^1 2_S$ and $^1 2_S$-P may explain why in the fully optimized complex H_2 can bind to the Fe atom but not to the Ni atom, whereas in the complex featuring the geometry observed in the enzyme H_2 can bind to the Ni atom but not to the Fe atom.

The analysis of the electronic properties of Ni-SI_a models suggests that the S_a–Ni–S_b angle is also crucial for tuning the reactivity of H_2. In fact, analysis of the population of the σ and σ^* orbitals of H_2 reveals that the former decreases and the latter increases as a function of the Sa–Ni–Sb angle (see figure 4.9), indicating that both donation of electron density from H_2 to Ni and back-donation from Ni to H_2 are affected by the value of this angle. Correspondingly, the H–H bond distance increases as a function of the Sa–Ni–Sb angle. The large increase of the σ^* orbital population suggests that H_2 could be activated through an oxidative addition mechanism.

After the analysis of the factors affecting the H_2 binding step, we have investigated plausible reaction pathways leading to H_2 cleavage. In particular, starting from the lowest energy (i.e., most favored) Ni–H_2 adduct (which corresponds to H_2 binding to the Ni ion in the low spin Ni-SIa redox state), we have considered two different pathways (figure 4.10): (A) heterolytic cleavage of H_2 in which the S atom of Cys555 extracts a proton from H_2, with the concomitant formation of a terminal hydride coordinated to Ni, followed by transfer of the hydride to a bridged position between the two metal atoms; (B) oxidative addition of H_2 with the formation of an intermediate featuring one terminal and one bridged H atom, followed by proton transfer from Ni to the S atom of Cys555. We have also considered analogous pathways in which the S atom of Cys61 is involved in the proton abstraction from H_2. However, activation and reaction energies were systematically higher (data not shown), indicating that Cys555 is the residue that is transiently protonated upon H_2 cleavage. The pathways featuring protonation of Cys61 may be less favorable due to the formation of two H-bonds between the sulfur atom of Cys61 and the amide hydrogen atoms of residues Val63 and Cys64, which make the sulfur

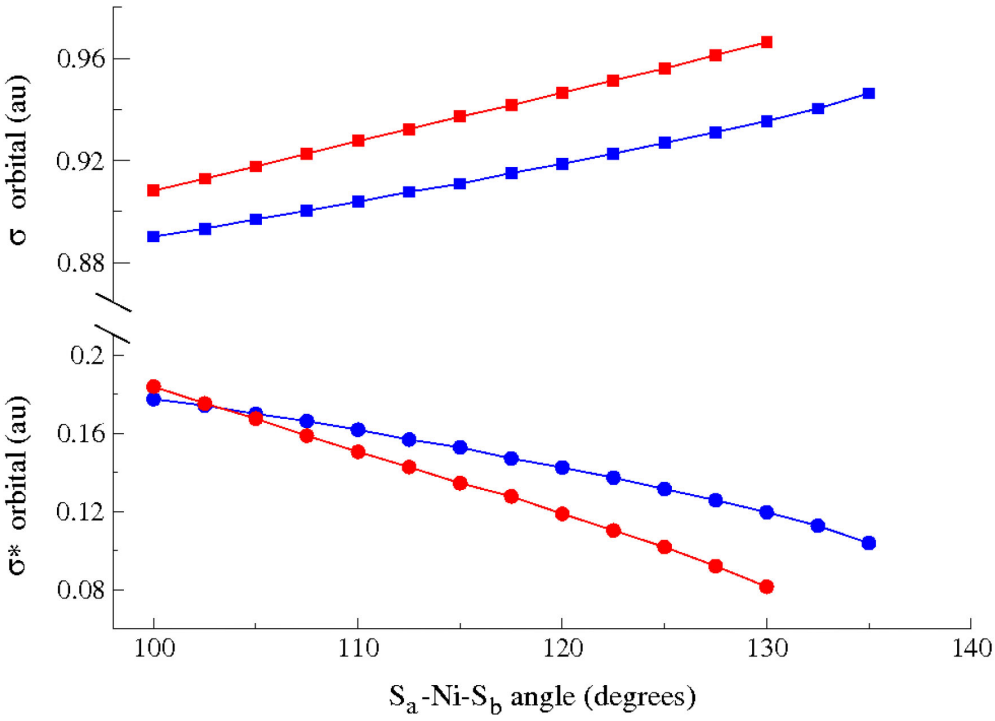

Fig. 4.9 NBO σorbital population of H_2as a function of the S_a-Ni-S_b angle in the complex ^{1}2$_S$-Ni$_{H2}$ (blue line with square symbols) and in the complex $[(CH_3S)_4(H_2)Ni]^{2-}$ (red line with square symbols). NBO σ* orbital population of H_2as a function of the S_a-Ni-S_b angle in the complex ^{1}2$_S$-Ni$_{H2}$ (blue line with circle symbols) and the complex $[(CH_3S)_4(H_2)Ni]^{2-}$ (red line with circle symbols). Only results in the range 100-135° are reported, because H_2does not bind to Ni for larger values of the angle.

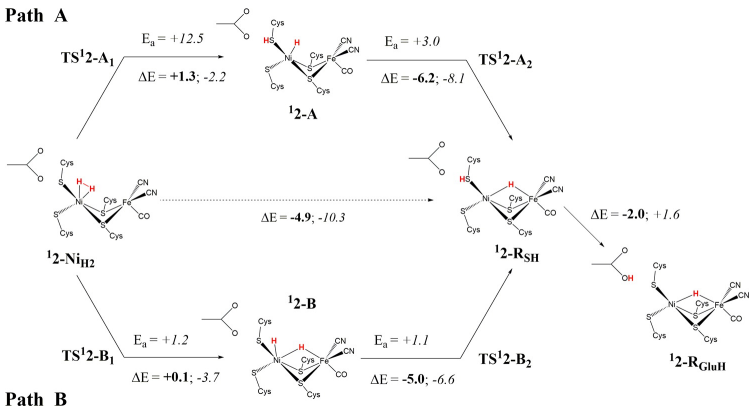

Fig. 4.10 Schematic Representation of the Possible Reaction Pathways for Dihydrogen Activation and Cleavage in the Active Site of [NiFe]-Hydrogenases. Values in bold refer to calculations performed using the large model, whereas values in italic refer to calculations performed using the medium-size model. Energy differences and activation energies are in kcal mol^{-1}

atom less basic, and provide some steric hindrance for its protonation.

In the following, the atoms of hydrogen which migrate to the S atom of Cys555 (S_c in 4.2) and to the NiFe bridging position will be labeled Hb and Ha, respectively. Geometrical parameters discussed in the text will refer to the large model for minimum energy species and to the medium-size model for transition states (TS). Comparison of selected geometry parameters of the large- and medium-size models is given in figure 4.11.

The first step in pathway A transiently yields a terminal hydride species (^{1}2-A in figure 4.10 and figure 4.11). A similar terminal hydride species was characterized for the more oxidized Ni-C form of the enzyme by Neese et al. [115], even if it was 22 kcal mol^{-1} less stable than the corresponding μ-hydride form. The activation energy for this step is 12.5 kcal mol^{-1}. In the corresponding transition state ($TS^1$2-A_1 in figure 4.10 and figure 4.11), the two atoms of dihydrogen are 1.28 Å apart. The intermediate ^{1}2-A is almost isoenergetic with the reactant ^{1}2-Ni_{H_2}; the energy difference is +1.3 and -2.2 kcal $mol^{-1}$for the large- and medium-size models, respectively. In ^{1}2$_L$-A, the Ni atom features a pseudo square pyramidal coordination, in which the S atom of Cys61 ($S_a$in figure 4.2) occupies the apical position. The Ha–Ni distance and the H_a–Ni–Fe angle

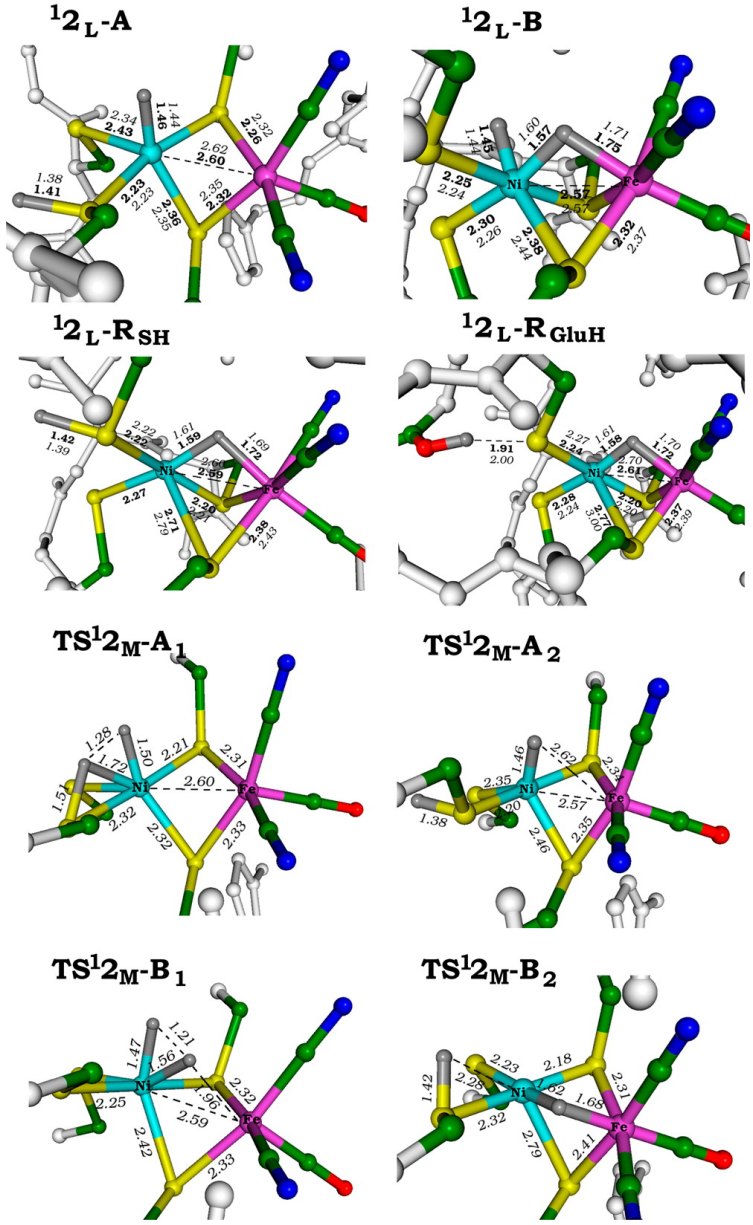

Fig. 4.11 Schematic representation of the optimized structures of the species in pathways A and B with selected distances in Å. In bold and italic are shown the distances calculated using the large- and medium-size models, respectively. Transition states have been calculated only using the medium-size model, and therefore, only one set of distances is reported. For the sake of clarity, hydrogen atoms, with the exception of H_a and H_b, are not shown.

in $^{1}2_{L}$-A are 1.46 Åand 100.3°, respectively.

$^{1}2$-A is a transient species that, in the second step of pathway A, can quickly rearrange to the μ-hydride species $^{1}2$-R_{SH}. The activation energy for this step is as small as 3.0 kcal mol^{-1}, and the reaction is exoenergetic by -6.2 and -8.1 kcal mol^{-1}considering the large- and medium-size models, respectively. The geometry of the transition state ($TS^{1}2\_A_2$ in figure 4.10 and figure 4.11) is similar to that of the intermediate, with the Ha–Ni and Ha–Fe distances and the Ha–Ni–Fe angle equal to 1.46 Å, 2.62 Å, and 75.8°, respectively.

The reaction product of the second step ($^{1}2$-R_{SH} in figure 4.2 and figure 4.11) should correspond to the reduced Ni-R form of the enzyme [37, 73, 154]. In this species, Ha asymmetrically bridges the two metal atoms (Ha–Ni = 1.59 Å; Ha–Fe = 1.72 Åin $^{1}2_{L}$-RSH), whereas Hb is migrated as a proton to the Sc atom. This form is characterized by a significant elongation of the Ni–Sb bond (2.7 Å$^{1}2_{L}$-R_{SH}). Pathway B is also characterized by two steps. In the first step, H_2 undergoes a formal oxidative addition, in which Hb remains terminally coordinated to Ni and Ha moves to a bridging position between the two metal atoms, yielding the $^{1}2$-B intermediate (see figure 4.10 and figure 4.11). The oxidative addition step in pathway B is almost barrierless (activation energy 1.2 kcal mol^{-1}). The intermediate $^{1}2$-B is isoenergetic with the reactant (ΔE= +0.1), or more stable than the reactant by -3.7 kcal mol^{-1}, for the large- and medium-size models, respectively. Also, the second step of pathway B, corresponding to the Ni-to-S_c proton migration, is almost barrierless (E_a= 1.1 kcal mol^{-1}), leading to the product $^{1}2_{L}$-R_{SH} which is -5.0 or -6.6 kcal mol^{-1}lower in energy than $^{1}2$-B for the large- and medium-size models, respectively.

In the transition state of the first step of path B ($TS^{1}2$-B_1 in figure 4.10 and 4.11), the H_a atom asymmetrically bridges the Ni and Fe atoms (Ha–Ni = 1.56 Å; Ha–Fe = 1.96 Å), while the Hb atom is moving to a terminal position on Ni (Hb–Ni = 1.47; Hb–Ni–Fe = 91.8°). The Ha–Hb distance and the Ni–Fe bond length are 1.21 and 2.59 Å, respectively. The intermediate $^{1}2_{L}$-B has a pseudo octahedral coordination of the Ni atom, in which the H_b–Ni–H_a angle is 77.2° and the Ha–Hb distance is 1.89 Å.

The bridging Ha atom is slightly closer to Ni than Fe (Ha–Ni = 1.57 Å; Ha–Fe = 1.75 Å), while the terminal $H_b$atom (H_b–Ni = 1.45 Å) is slightly bent in the direction of S_c(H_b–Ni–S_c= 82.9°). The H_b atom moves to an asymmetric bridging position between Ni and Sc in the transition state of the second step of pathway B ($TS^12\_B_2$ in figure 4.10 and figure 4.11), in which it is much closer to S_c than to Ni (H_b–S_c = 1.42 Å; H_b–Ni = 2.28 Å). The Ni–$S_b$distance in TS^12-B_2 is 2.79 Å, and is very similar to that observed in the product ^12_L-RSH.

As shown in figure 4.10, we have also investigated the proton transfer from the Sc atom of Cys555 in ^12_L-R to the carboxylate group of Glu14 (^12_L-RGluH in figure 4.10 and figure 4.3). This process is almost isoenergetic (ΔE = -2.0 and +1.6 kcal $mol^{-1}$for the large- and medium-size models, respectively), suggesting that ^{12}-R_{SH} and ^{12}-R_{GluH} are in equilibrium, and may correspond to two different Ni-R forms characterized by spectroscopic methods, as also recently suggested by Neese et al. [71].

It is interesting to note that, after the monoelectronic oxidation of ^12_L-RSH yielding the putative Ni-C form of the enzyme, the species in which Cys555 is protonated does not correspond to an energy minimum, because the proton spontaneously migrates during geometry optimization to the carboxylic group of Glu14. In addition, the Ni–Sb bond length significantly shortens to the value of 2.40 Å, in very good agreement with the distance observed in the crystallographic structures of reduced forms of the enzyme (data not shown) [82, 103].

Previous theoretical investigations suggested that H_2 binding and/or cleavage can occur on species where the redox state of Ni is more reduced ($NiII^I$) or more oxidized (Ni^{III}) than Ni^{II}. Therefore, with the aim of obtaining homogeneous and comparable results, we reinvestigated the reaction profiles of these mechanisms using our largest model.

A mechanism involving the NiI redox state in H_2 binding and cleavage has been recently proposed [131]. According to this scenario, heterolytic cleavage of H_2 takes place on Ni-SIa only in the initial step. Then, a NiI form of the active site is formed, which was proposed to efficiently bind and cleave H_2 molecules via an oxidative addition step. In order to reinvestigate this mechanism, we started from a $NiII^I$ form (hereafter

labeled $^{2}1_{L}$) obtained by one-electron reduction of $^{1}2_{L}$. H_2 can bind to the Ni atom of $^{2}1_{L}$, yielding the H_2 adduct $^{2}1_{L}$-NiH_2 (see figure 4.12), but the process is endoenergetic by 2.5 kcal mol^{-1}. The corresponding adduct in which H_2 binds to the Fe^{II} ion of $^{2}1_{L}$ is 20.1 kcal mol^{-1}higher in energy than the reactants, and its formation can therefore be ruled out. In $^{2}1_{L}$-NiH_2, the Ni–H bonds are longer (1.701 and 1.710 Å) than in $^{1}2_{L}$-NiH_2, and correspondingly, the H_2 molecule is less activated. Oxidative addition of H_2 on the Ni atom of $^{2}1_{L}$-NiH_2 yields the intermediate $^{2}1_{L}$-B (see figure 4.12). However, $^{2}1_{L}$-B is higher in energy than the H_2 adduct by about 7 kcal mol^{-1}, a value which added to the H_2 binding energy indicates that this reaction pathway is energetically disfavored when compared to H_2 activation and cleavage on a Ni^{II} species. In addition, the step corresponding to the proton transfer from Ni to Cys555 to give $^{2}1_{L}$-X* is also endoenergetic by about 4 kcal mol^{-1}. The intermediate in which H_2 is cleaved heterolytically with protonation of Cys555 and concomitant formation of a terminal hydride coordinated to Ni has not been characterized, because the terminal hydride spontaneously moves to the bridging position during geometry optimization. Therefore, the heterolytic cleavage of H_2 on the $NiII^{I}$ form yields in one step the $^{2}1_{L}$-X* species, which is about 11 kcal mol^{-1}higher in energy than $^{2}1_{L}$-NiH_2.

Another scenario was explored by Hall et al., showing that H_2 cleavage on a Ni^{III} form of the active site is more favorable than on the corresponding high and low spin Ni^{II} forms, suggesting that binding of H_2 on Ni^{II} is followed by oxidation to a $Ni^{III}H_2$ adduct [117, 149, 159]. In fact, according to the calculations carried out using our large model, H_2 binding can occur to both Ni^{III} and Fe^{II} ($^{2}3_{L}$-NiH_2 and $^{2}3_{L}$-FeH_2), and both H_2 adducts are almost isoenergetic with the reactants (ΔEH_2Ni= -1.0 kcal mol^{-1}; ΔEH_2Fe = -0.1 kcal mol^{-1}). A formal oxidative addition reaction on $^{2}3_{L}$-NiH_2, to yield an intermediate featuring one terminal and one bridged H atom ($^{2}3_{L}$-B), is an endoenergetic process by 2.5 kcal mol^{-1}(see figure 4.12). On the other hand, the heterolytic cleavage yielding an intermediate ($^{2}3_{L}$-A) featuring one proton on Sc and one hydride terminally coordinated to the Ni atom is strongly exoenergetic (-17.1 kcal mol^{-1}). It should be noted that in $^{2}3_{L}$-A the proton initially placed on Sc spon-

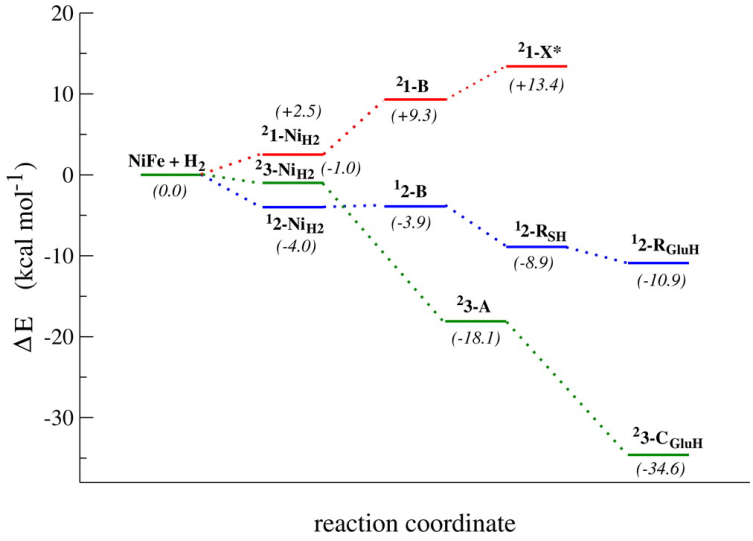

Fig. 4.12 Energy profile of H_2binding and cleavage along the lowest energy pathway (only minimum energy species are considered) for the low spin Ni^{II} (blue line), $NiII^{I}$ (red line), and Ni^{III} redox forms of the active site. In parentheses are the relative energies (in kcal mol^{-1}) with respect to NiFe + H_2 (NiFe = ^{2}1, ^{1}2, or ^{2}3) using the large model.

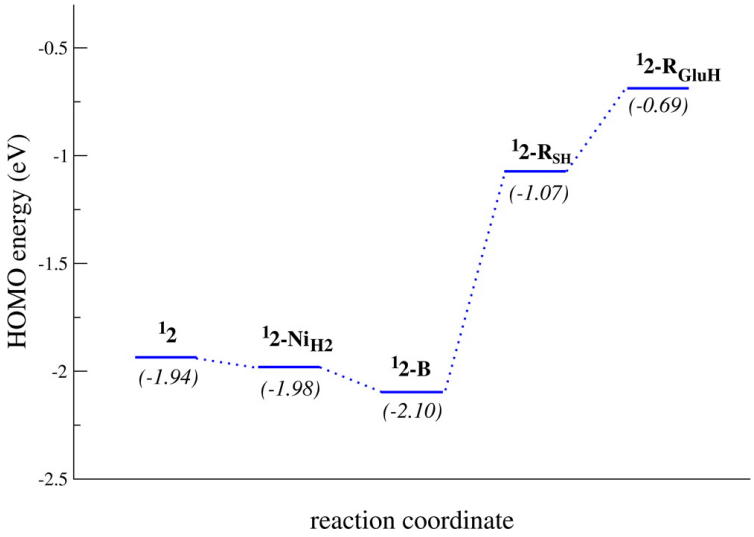

Fig. 4.13 HOMO energies along the lowest energy pathway (only minimum energy species are considered) for the low spin Ni^{II}redox form of the active site. In parentheses are the HOMO energies in eV.

taneously transfers to Glu14 during geometry optimization. Migration of the hydride from the terminal to the bridged position, ^{2}3-C_{GluH} is also an exoenergetic step by -16.5 kcal mol^{-1}. Therefore, cleavage of H_2 on the Ni^{III} form of the active state is calculated to be an exoenergetic process by about 35 kcal mol^{-1}. The large stabilization of the ^{2}3-C_{GluH} product when compared to the H_2 adduct may suggest that, indeed, oxidation of Ni^{II} to Ni^{III} and H_2 cleavage might occur through a concerted mechanism. In this respect, a rough estimate of the oxidation propensity of the active site can be obtained analyzing the HOMO energies of the species involved in the reaction mechanism. Figure 4.13 shows the HOMO energies of the species along the lowest energy profile in the cleavage of H_2 on the low spin Ni^{II} form. It is interesting to note that the HOMO energy is nearly constant along the reaction pathway in ^{1}2, ^{1}2_Ni_{H_2} and ^{1}2_B, whereas it increases significantly in the ^{1}2-R_{SH} product, suggesting that an electron can be easily extracted from the latter form. This observation is in line with the experimental evidence that Ni-$SI_a$and Ni-R forms in the enzyme are in a non-redox equilibrium.

Another plausible mechanism might start with the H_2 cleavage on the Ni^{II} form but then proceed by a double proton and electron loss to form a $NiII^0$ species that can bind and cleave H_2 in an oxidative addition step. However, geometry optimization of such a $NiII^0$ form (hereafter labeled $^1 0_L$) obtained by two-electron reduction of $^1 2_L$ yields to a significant reorganization of the active site, with the S atom of Cys558 moving at a distance of about 3.2 Å from Ni and one CN^- ligand migrating to a semibridged position between the two metal atoms. In addition, binding of H_2 to Ni yields to a species (^{1}0-Ni_{H_2}) featuring a full dissociation of the S atom of Cys555 and a significant rearrangement of the ligand configurations around the Ni atom, whereas binding of H_2 to Fe leads to an adduct (^{1}0-Fe_{H_2}) which is more than 20 kcal mol^{-1} higher in energy than the reactants. Finally, we also investigated the mechanism in which H_2 might bind to Ni^{II} in an enzyme form in which a hydride is already bridged between the two metal atoms. However, in this case, all attempts to obtain a H_2 adduct resulted in the dissociation of H_2. In addition, a plausible intermediate in which H_2 is heterolytically split in a terminal hydride on Ni and a proton on S_c evolves during geometry optimization to a structure in which the S atom of Cys61 moves at a distance of about 4 Å from Ni, leaving the metal atom in square pyramidal configuration with the bridging hydride and the terminal hydride occupying the apical and an equatorial position, respectively.

4.5 Conclusions

The disclosure of the stereoelectronic properties of the active site of [NiFe]-hydrogenases which are crucial for efficient H_2 binding and cleavage is important not only in the context of the efforts aimed at elucidating structure–function relationships in hydrogenases [76, 195] but also for the rational design of novel bioinspired catalysts for H_2 oxidation [41, 42, 78, 152, 153, 183]. To contribute to such an effort, we have used DFT to study the H_2 binding and activation properties of models of the active site of [NiFe]-hydrogenase, ranging from minimal (metal ions and first coordination sphere) to very large (metal cluster, first and second

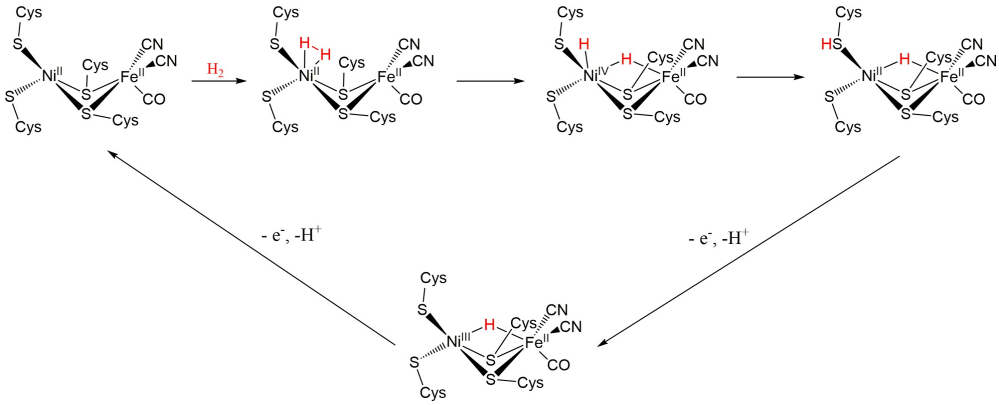

Fig. 4.14 The Most Plausible Catalytic Mechanism for H_2Oxidation in [NiFe]-Hydrogenases, according to the Computational Results Obtained in the Present Work.

coordination sphere – 290 atoms). Results highlight that the spin state and the coordination geometry of the Ni ion are two crucial factors that tune the energetics and regiochemistry of H_2 binding to the NiFe metal cluster. In particular, (a) H_2 binding is energetically favored only when the NiFe metal cluster is in a low spin state and the Ni cysteine ligands have a distorted seesaw coordination geometry, as observed in the crystallographic structures of [NiFe]-hydrogenases. In addition, (b) when the Ni coordination sphere has such peculiar stereoelectronic features, H_2 binding is regiospecific and takes place exclusively at the Ni center. Two low activation energy pathways are possible after H_2 binding to the NiII ion: direct heterolytic cleavage of H_2 mediated by the sulfur atom of one of the terminally coordinated cysteine residues or oxidative addition of H_2 followed by proton transfer to the sulfur atom of one of the terminally coordinated cysteines. The latter reaction pathway is almost barrierless, indicating that H_2 cleavage can occur very quickly leading to the putative Ni-R form (figure 4.14).

The oxidative addition pathway also reveals that a transient (formal) Ni^{IV}species could be formed in the catalytic cycle of [NiFe]-hydrogenases. Notably, the product of H_2 cleavage formed according to an oxidative addition step followed by proton transfer to the S atom is indistinguishable

from that obtained from direct heterolytic cleavage, and therefore in full agreement with experimental observations indicating heterolytic cleavage of H_2 in the active site of [NiFe]-hydrogenases [154]. More importantly, our results show that activation and cleavage of H_2 is strongly affected by the geometry of the ligands coordinated to the Ni atom, which must have a distorted seesaw arrangement (S_b–Ni–S_a< 135°; S_c–Ni–S_d> 160°) for efficient catalysis.

The formal transient formation of a Ni^{IV}dihydride species into the enzyme active site was never proposed before, but it is not without precedent when considering the chemistry of Ni-containing synthetic catalysts. In fact, DuBois et al., which have characterized the reactivity of several nickel complexes containing proton relays [189], reported the detailed experimental dissection of the mechanism for electrochemical H_2 oxidation catalyzed by $[Ni(dppp)(PPH_2NBz_2)](BF_4)_2$, where PPH_2NBz_2 is 1,5-dibenzyl-3,7-diphenyl-1,5-diaza-3,7-diphosphacyclooctane and dppp is 1,3-bis(diphenylphosphino)propane [221]. Similarly to our observations, DuBois et al. found that the initial H_2 addition step involves the formation of a transient dihydrogen complex, which then undergoes a rate determining H–H bond cleavage to form the corresponding Ni^{IV} dihydride complex, for which spectroscopic evidence was obtained. Then, the dihydride species evolves according to a reaction step in which one of the two hydrogen atoms coordinated to Ni is transferred as a proton to the N atom of the cyclic ligand. Notably, in the computed structure of the transient dihydrogen complex $[Ni(dppp)(PPH_2NBz_2)]^{2+}$, the phosphine ligands do not remain in a slightly distorted square planar configuration (as observed in the reactant $[Ni(dppp)(PPH_2NBz_2)]^{2+}$) but assume an arrangement similar to the seesaw geometry observed in the enzyme.

The observation that a seesaw Ni coordination geometry is a key feature for efficient H_2 binding and activation can be usefully discussed also in the light of other previous results that showed how H_2 binding and activation on Ni-containing functional models of hydrogenases is often concomitant with a drastic structural rearrangement of the metal coordination geometry. In particular, it was previously reported that H_2 binding and cleavage on $[Ni(NHPnPr_3)(s_3)]$ (where s_3 = bis(2-sulfanylphenyl)sulfide)

is concomitant with the rearrangement of the Ni coordination environment from a square planar to a flattened tetrahedral geometry [223]. However, in the case of $[Ni(NHPnPre_3)(s_3)]$, it was found that such a structural rearrangement is energetically very disfavored and this is the main cause of its poor catalytic properties. Our results show that such rearrangement does not take place in the enzyme, because in the four-coordinated low spin Ni species corresponding to the Ni-SI_a enzyme form, due to the stereoelectronic properties of the metal cluster environment, the Ni coordination sphere already has a distorted seesaw geometry, and not the classical square planar geometry that would be expected for simpler low spin Ni^{II} coordination compounds.

From an electronic structure perspective, our results complement the observations made by DuBois et al., which highlighted how an increased driving force for H_2 addition is linked to the distortion of square planar Ni complexes, essentially because such a distortion leads to a decrease in the energy of the LUMO, which enhances the hydride acceptor ability of NiII complexes [62, 168]. In addition, the observation that reaction energies for H_2 binding and cleavage in the minimal model $[(CH_3S)Ni]^{2^{\smile}}$ are systematically less favorable than in bimetallic active site models underlines the noninnocent role of the $Fe(CO)(CN)_2$ moiety, and might encourage further studies aimed at understanding how the communication between the metals facilitates the electron transfer processes throughout the catalytic cycle [216].

Chapter 5

Active site protein dynamics in phosphotriesterase

5.1 Introduction

The study of structure-dynamic-function relationship in proteins has revealed some of the fundamental principles of the mechanisms related to protein activity, function and interactions [99, 116, 121]. Indeed, in the last years it has been enforced the notion that protein can be best described as dynamical entities and that the protein native state can be populated by several different substates in equilibrium each other [106, 171]. Dynamics can play a key role in the transitions among those substates, which are often related to different functions or recognition properties [20, 30, 169, 171, 203]. Methods inspired by graph theory are often employed to detect paths of communication between distal sites in a protein in the attempt to provide an atomistic picture of the residues which are likely to be involved in the propagation of structural effects and promotion of conformational changes [17, 40, 48, 70, 90, 175, 207]. Related to this, the propagation of the structural effects over long distances has been suggested to occur by multiple and pre-existing paths of communication, which are embedded in the network of residue contacts and related to the intrinsic dynamics of the proteins [56, 141]. The study in atomistic details of those mechanisms is of fundamental relevance, for example, to

rationalize the effects induced by known distal and secondary mutations on the functional sites [31, 79, 80, 111, 201, 218].

In this context, Molecular Dynamics (MD) simulations can help to disclose protein dynamics and to predict the effect induced by mutations [19, 22, 38, 43, 46, 50, 80, 88, 125, 126, 136, 137, 158, 178, 180, 196, 215, 219]. In fact, even if limited by the achievable sampling and force field accuracy, recent studies in the field attested the potential of MD in providing information on a wide range of molecular motions [58, 61]. Protein dynamics on different timescales can be linked as showed by NMR investigations [99], implying that the motions on ps or ns timescales may potentially provide information on the events that modulate important conformational transition and occur at longer timescales. Also, cascade effects inducing changes in side-chain conformations starting from a remote site have been shown to perturb the active site residues of a proline isomerase, and mutations, which were designed to alter the population of the conformational states of the protein, were able to also affect the catalytic properties [67, 79]. These cascade effects can be described as networks of residues in communication, and several of these networks have been recently identified by contact analyses and sampling of the electron density of 35 X-ray structures of different proteins [201]. Moreover, methods inspired by graph theory have been often coupled to MD simulations to study paths of long-range communication within proteins or protein complexes [89, 126, 137, 155, 158, 174, 176]. It is still not clear, although, how many of the paths that these approaches identify are then also biologically-relevant and further studies are needed in the field toward a more conscious usage of the information that the protein structure network can provide. We here turned our attention to the phosphotriesterases (PTEs, also known as organophosphorus-hydrolases) isolated from *Agrobacterium radiobacter* (arPTE, also know as OpdA), a metalloenzyme that is generally active over a large variety of metal cluster in both homonuclear and heteronuclear complexes [26, 29, 68, 77]. Nevertheless, most of the experimental characterizations of PTE enzymes were carried out in presence of Zn^{2+} at the active site. PTE enzymes have a general remarkable interest since they are active on a broad spectrum of

organophosphate triesteres [142], making them attractive for applicative purposes in biodegradation of organophosphate pesticides, which account for about 36% of the total world market. Gene shuffling experiments on arPTE allowed to identify variants of the enzyme bearing multiple variants with increased or decreased activity on a methylparaoxon substrate with respect to wild type arPTE [110]. Most of the mutation sites were placed at remote sites with respect to the catalytic residues. In particular, the so-called arPTE-4M (K185R, D208G, N265D, T274N) and arPTE-8M (G60A, A80V, R118Q, K185R, Q206P, D208G, I260T, G273S) mutants emerged from this study. Not only the kinetic parameters of arPTE-4M and 8M variants have been experimentally characterized, but also their three-dimensional (3D) structures were solved by X-ray crystallography [110]. Indeed, the seminal work from Jackson and coworkers [110] explored the conformational landscape of arPTE from the interpretation of the electron density maps obtained by X-ray crystallography. arPTE conformational landscape has been described as being characterized by two dominant conformations: one, E_{closed}, that is pre-organized for substrate hydrolysis, and the other one, E_{open}, that favors substrate and product diffusion. While the two conformations are overall quite similar, they feature differences close to the active site, notably in residues that are in proximity of or belonging to loop 4 (Phe132, Ala171, Thr172, Thr173, Gly174 and Lys175) and 7 (Arg254, Try257, Leu 271 and Phe272) (Figure 5.1). In particular, in the proximity of loop 7 is located, some changes can be observed when the E_{closed} state is compared to the E_{open} one (Figure 5.1). In particular, Arg254 is bent away from the active site, Tyr257 is tilted in the same direction, and the adjacent loop 7 overhanging the active site is also in a outward position with respect to the active site, resulting in the movement of Leu271 and Phe272. On the opposite side of the catalytic cleft, where loop 4 is located, Phe132 flips away from he active site in E_{open}, in concert with a shift of loop 4 (residues 171-175). Jackson et al. [110] show that a correlation exists between the reaction rate and the population of the E_{open} or E_{closed} states. Indeed, the two arPTE mutants, 4M and 8M feature increased and reduced k_{cat} if compared to the wild-type variant [110], respectively. Indeed, the elec-

tron density map of wild-type arPTE shows a higher occupancy (0.7 for the E_{closed} state. arPTE-4M, which has a higher k_{cat} than the wild-type variant, presents both closed and open conformations in equilibrium (occupancy values in the X-ray structure of 0.4 and 0.6 for E_{closed} and E_{open}, respectively) [110]. In particular, arPTE-4M has an increased occupancy of E_{open} with respect to the wild-type enzyme. Instead, E_{open} is the dominant conformation of the arPTE 8M variant (1.0 of occupancy), which features reduced k_{cat} respect to the wild-type enzyme, meaning that the enzyme cannot recover the closed conformation after substrate binding, thus diminishing its reaction rate.

Overall, these data thus suggest that remote mutations can affect the turnover rate of arPTE by altering its conformational landscape. The long-range communication between the mutation sites and the residues in the proximity of the active site can thus play a role. We here used arPTE as a model system to verify the robustness and capability of PSN methods coupled to MD [89, 156] to detect paths between the known mutation sites and the residues in the proximity of the loops 4 and 7, which are the sites where the conformational changes between "open" and "closed" states occur. Indeed, the availability of the experimental structures for the two conformations and the knowledge of their relationship to protein activity provides us with a valuable source of information to assess the PSN/MD method and the different parameters involved, as well as to provide atomistic details on the residues involved in the conformational changes and the paths that they mediate, enriching the knowledge that we have on this system.

5.2 Materials and Methods

5.2.1 Molecular dynamics simulations

The known X-ray 3D structures of arPTE wild-type (PDB entry 2R1N) and 8M (pdb entry 3A3W) were used as starting structures for 50 ns MD simulations with a modified version of the CHARMM22* force fields ([163]). The original CHARMM force field was supplemented with param-

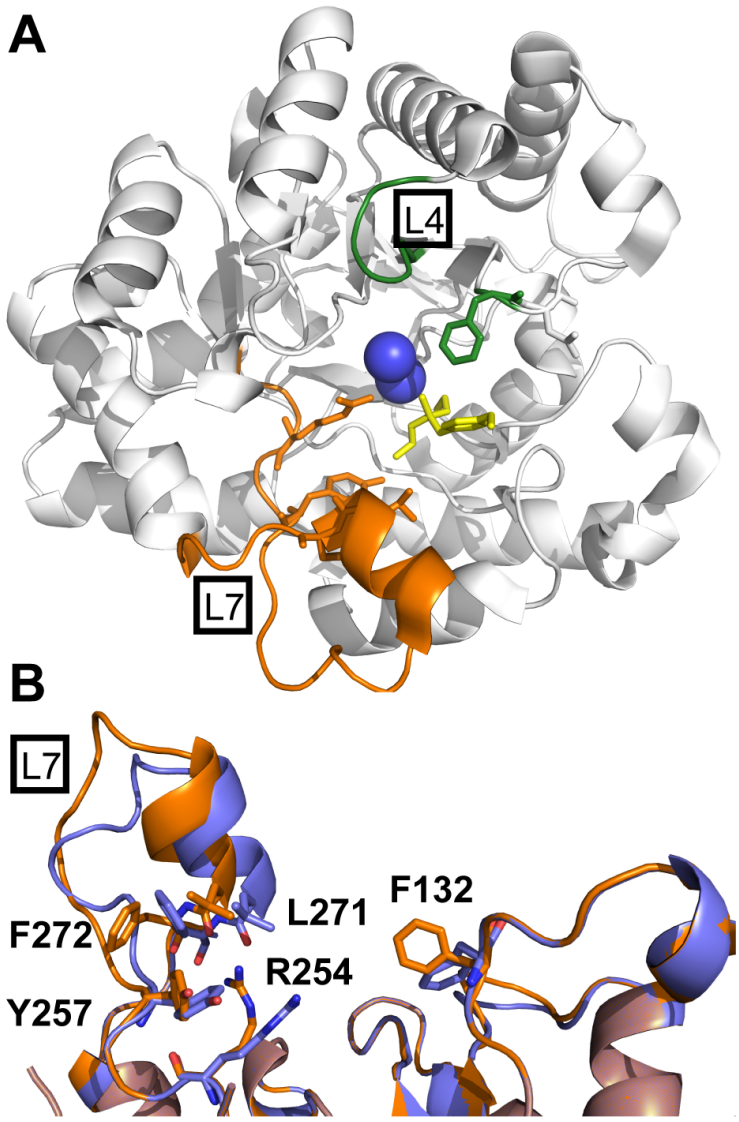

Fig. 5.1 3D structure of arPTE. A) Structure of arPTE. Loop 7 and Loop 4 regions are colored in orange and green, respectively. The substrate 4-methoxylphenyl diethyl phosphate substrate is represented with cyan sticks. The two metal atoms bound at the catalytic site are shown as blue spheres; the residues shown as sticks are those involved in the conformational change between E_{open} and E_{closed} states. B) Residues involved in the open/closed conformational transition are shown in sticks. E_{open} and E_{closed} conformations are shown as orange and blue cartoon, respectively. Both conformations are present in the electron density of the 4M mutant, which is here used to showcase the open/closed conformations (PDB ID 3A4J). The residues involved in the conformational change between the two states are depicted as sticks.

eters from the ParamChem web server [202] in order to supplement parameters for the terminal carbamic acid group of the carboxylysin (KCX) residue. To reduce artifacts from lack of convergence, 3 independent MD simulations of each variants were carried out, as well as one simulation per system was extended to 100 ns. Simulations were carried out in explicit solvent using the 4.5.5 version of the GROMACS (GROningen Machine for Chemical Simulation) software [101]. The initial structures were embedded in a dodecahedral box of 9420 and 8990 TIP3P water molecules respectively arPTE-WT and arPTE-8M, and periodic boundary conditions were employed. The following preparation steps have been followed for both the arPTE-WT and the arPTE-8M solvated systems. Each system was initially relaxed by minimization (steepest descent, 50.000 steps). The optimization step was followed by 100 ps of solvent equilibration at 300 K, while restraining the protein and ions atomic positions using an harmonic potential. Following this step, the system was slowly heated to the target temperature (298 K) and pressure (1 bar) through an unrestrained NPT MD simulation. From this simulation, the trajectory frame having a volume as close as possible to the average volume of the simulation was selected, and was used simulation as the starting structure for an NVT performed Productive 100 ns MD simulations were carried out in the isothermal-isobaric (NPT) ensemble, using an external Berendsen thermostat with a coupling constant of 0.1 ps at 300 K. Productive MD simulations were performed in the NVT ensemble at 298K and 1 bar, using the thermostat of Bussi et al. with thermal coupling of 0.01 ps. The LINCS algorithm[100] was used to constrain heavy-atom bonds, allowing for a 2 fs time-step. Long-range electrostatic interactions were calculated using the Particle-Mesh Ewald (PME) summation scheme[51]. Van der Waals and Coulomb interactions were truncated at 0.9 nm. The non-bonded pair list was updated every 10 steps and conformations were stored every 4 ps.

5.2.2 Analysis of MD simulations

The main chain root mean square deviation (rmsd), which is a crucial parameter to evaluate the stability of MD trajectories, was computed us-

ing the corresponding X-ray structure as a reference. The first 5 ns of each trajectory were discarded as initial equilibration. DSSP program was used to calculate secondary structure content[113], along with a residue-dependent persistence degree of secondary-structure profile (pdssp) to assess stability of secondary structures. The distance between the Zn^{2+} and its ligands has been checked throughout the simulations, and was found to be remarkably stable, compatibly with consistent ion coordination at the active site.

5.2.3 Evaluation of the MD ensembles

At first, the main-chain root mean square deviation with respect to the starting structures for simulations was evaluated over the simulation frames (Figure 5.2 and 5.3). 5 ns are generally required to ensure a plateau in the rmsd and are thus discarded from further analysis. Since we collected independent simulations for each system, we evaluate if each of them was consistently described the dynamics of the different variants. In particular, we used the root mean square inner product of the first 20 eigenvectors from the Principal Component Analysis applied to the MD ensemble both comparing pairwise the independent replicates of the same system and the two halves of the same trajectory (Figure 5.4). It turns out a reasonable overlap with rmsip values higher than 0.63 in all the cases, indicating a robust set of simulations for each system.

To further assess the capability of the collected simulation to give a congruent view of protein dynamics, we also calculate the per-residue root mean square fluctuations (rmsf) of Cα atoms at using a time-window of 5 ns as an index of protein flexibility and then we compared each rmsf profiles of the same system using the Pearson correlation coefficient. The flexibility profiles achieved for each independent replicates of the same variant were in high agreement with correlation values higher than 0.9. At last, if each individual trajectory is consistently describe the dynamics of the variant of interest the dynamics of the side-chains should be overall conserved in each replicate of the same variant. As a consequence, the distribution of the side-chain dihedrals should be comparable in independent simulation of the same system. We thus calculated the χ_1

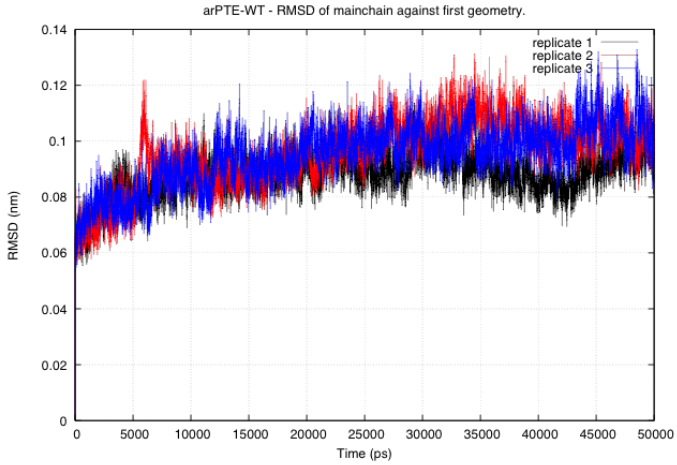

Fig. 5.2 Mainchian RMSD of the highly structured regions of the protein respect to the first frame, for the three replicates of arPTE-WT

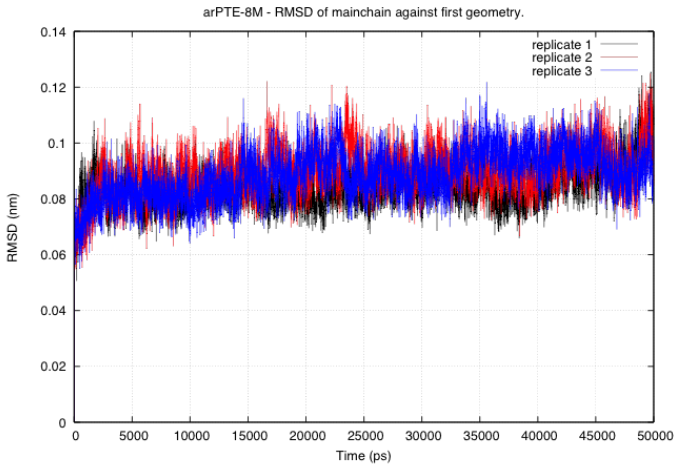

Fig. 5.3 Mainchian RMSD of the highly structured regions of the protein respect to the first frame, for the three replicates of arPTE-8M

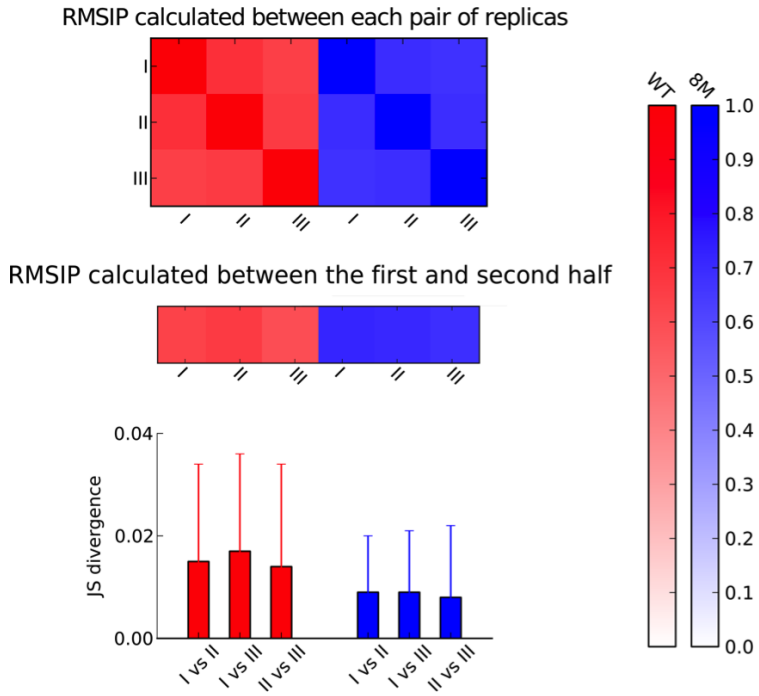

Fig. 5.4 Indeces of simulation stability for the MD simulations of arPTE-WT and arPTE-8M. Top: RMSIP values calculated between the first 20 eigenvectors of the essential subspaces of the MD replicates. Center: RMSIP values calculated between the 20 eigenvectors of the essential subspaces calculated on the first half and on the second half of each simulation. Bottom: average Jensen-Shannon divergence calculated between the distributions of $\chi 1$ and $\chi 2$ dihedrals calculated from each replicate.

and χ_2 dihedrals in each replicates and compared their distribution by the Jensen-Shannon divergence (d_{JS}) measure for each residue type. We then averaged all the d_{JS} obtained between each simulation pair, as shown in Figure 5.4. Indeed, the d_{JS} value is very low for the different replicates, meaning that the distributions of dihedral values are very similar.

5.2.4 Linear mutual information

The Linear Mutual Information (LMI) is a measure of motion correlation that, unlike cross-correlation, has the advantage of not depending on the relative orientations of the fluctuations, making it possible to identify correlated motions despite the difference between their orientations in space. LMI is defined in the [0;1] range, with 0 meaning no correlation and 1 meaning full motion correlation, without the distinction between anti-correlated and correlated motions. Given the joint probability distribution $p(\mathbf{r}) = Z/e^{-\beta U(\mathbf{r})}$, where Z is the partition function, $\beta = 1/T$ and $U(\mathbf{r})$ is the potential energy, and the marginal distributions $p_i(\mathbf{d_i}) = \int p(\mathbf{d}) d\mathbf{d}_i \neq j$, the relation

$$p(\mathbf{d}) = \prod_{i=1}^{3N} p_i(\mathbf{d_i}) \tag{5.1}$$

holds true if and only if the components d are uncorrelated. Violations of this equation due to possible correlations are quantified by the mutual information (MI), defined as

$$I(\mathbf{d_1}, ..., \mathbf{d_i}, ..., \mathbf{d_{3N}}) = \sum_{i=1}^{3N} H(\mathbf{d_i}) - H(\mathbf{d}) \tag{5.2}$$

where $H(\mathbf{d} = -\int p(\mathbf{d})) \log p(\mathbf{d}) d\mathbf{d}$ (i.e. the Shannon information entropy). The calculation of MI thus requires highly accurate and computationally demanding density estimates (p(d)); on the other hand, the density used in the less-demanding linear mutual information (LMI) is defined on a Gaussian approximation, employing the computationally efficient calculation of the covariance matrix. The used gaussian functions represent the quasi-harmonic approximation to the canonical density of

atomic motion, meaning that the calculated MI (LMI) contains only linear correlations. LMI between atoms i and j is calculated as:

$$I_{lin}(d_i, d_j) = 1/2(\ln|C_{(i)}| + \ln|C_{(j)}| + \ln|C_{(ij)}|) \tag{5.3}$$

where $C_{ij} = \langle(\mathbf{x_i}, \mathbf{x_j})^T(\mathbf{x_i}, \mathbf{x_j})\rangle$ and $C_i = \langle x_i^T x_i\rangle$.

LMI Correlation plots were obtained computing from the MD trajectories LMI values between Cα or Zn^{2+}atoms, using averaged overlapping windows of 2 ns, or 1 ns for the LMI network analysis.

5.2.5 Protein structure network

The PSN method employs the graphs formalism to define a network of interacting residues in a given protein or protein complex from the number of non-covalent interacting atoms, as has been explained more in details in the Results section. Briefly, a protein structure network (PSN) is a graph representation of the side-chain interactions between protein residues. Each residue is considered as graph node, and two nodes i and j are connected by an edge having weight equal to a calculated interaction strength (I) value:

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_i N_j}} \tag{5.4}$$

Where n_{ij} is the number of the distinct atom pairs between residues i and j at a distance lower than 4.5 Å. N_i and N_j are normalization values for residue types i and j which account for the different interaction propensities due to larger or smaller side-chains. A PSN edge is retained only if its interaction strength is higher than a prescribed cut-off (I_{min}). Similarly as what has been done previously for other moieties, and because of their important structural role, we have added the carboxylisine and residues to the PSN analysis. This required the calculation of the Ni normalization factor for both of them, which was performed by averaging the number of atomic contacts within 4.5 Å performed by both of them in 3D structures of Zn-binding PTE (PDB IDs 2OQL, 2O4Q, 2OB3, 1HZY, 1EYW, 1EZ2) The calculated values were 17.3 for Zn^{2+} and 58.2

for carboxylysine. We have used the PSN approach to calculate the most recurring shortest paths of communication in our simulation trajectories. For each simulation frame, a PSN at I_{min} 3.0, 4.0, 5.0, or 6.0 was calculated. Then, for every pair of residues of interest, the Floyd-Warshal algorithm was used in order to determine the shortest paths between selected pairs of nodes in the PSN graphs. Only the shortest paths in which at least one identified node featured a significant LMI correlation value of 0.4 or above with one of the residues of the selected pair was retained. Data from this analysis were gathered and only paths found in at least the 15% of frames were considered.

5.2.6 Network analysis of the LMI matrices

The network analysis of the LMI values for the Glycine mutation sites in the wild-type was performed on the graphs encoded in the 1-ns averaged motion correlation (LMI) matrices. Most importantly, we also included in the calculation the correlated motions between the protein residues and the Zn^{2+} ions, which are structurally important features of the enzyme, We have first filtered the LMI matrices by zeroing out all the matrix elements lower than 0.4, thus removing the least meaningful correlations. We have then considered the resulting filtered matrices as adjacency matrix of graphs, in which nodes are protein residues and edges have weight equal to the corresponding value in the LMI matrix. Finally, a shortest-path calculation was carried out on the obtained graphs between the Gly60 and Gly273 residues and those involved in the E_{open}/E_{closed} conformational transition, employing the graph analysis module of the PyInteraph package [199]. The obtained paths were plotted on the 3D structure using the xPyder ([161]) PyMOL plugin.

5.2.7 Analysis of protein interactions

The analysis of the hydrogen bonds and salt-bridge interactions was performed using the PyInteraph package, using the default cut-off values. The identified interactions were plotted on the 3D structure of proteins using the Interaction plotter PyMOL plugin, also part of the PyInteraph

package. In order to obtain a clearer representation, all the interactions with persistence lower than 5% were discarded.

5.3 Results

We here carried out three independent 50 ns MD simulations for wild type and 8M arPTE variants. In particular, we considered the wt structure used as a reference in Jackson et al. [110] and 8M variant in the E_{open} conformation. Before carrying out the path analysis, we evaluated the MD ensembles to ensure that a sufficient sampling was achieved (see Materials and Methods for details). We decided to run short different replicates of the same system instead that one longer MD simulation to avoid the possibility to encounter issues related to force-field deterioration of the protein structure and to collect an ensemble of structures closer to the experimental one. This also allows us to assess reproducibility of the results on the three different replicates. One of the best performing force field so far described and validated against experimental data is here employed, CHARMM22* [157]. We here employed a PSN-MD approach [89, 156] to detect paths of long-range communication in the protein. As stated in the Introduction, this method is based on the fact that structural effects can be transmitted at distal sites through communication paths involving side-chain contacts between residues. In this scenario, non-bonded Van der Waals interactions, which dominate protein structures, are the main component of the structural communication [207]. The approach relies on two main steps. i) the calculation of the PSN from the overall MD ensemble; in this network the residues are the nodes of the graph and are connected by edges weighted according to a defined "interaction strength" (I) value [34]. I between two residues is calculated from the number of interatomic contacts between their side-chains. In particular, a contact between two atoms belonging to two different residues is generally accounted when the two atoms are at a distance lower than 0.45 nm [34].. Finally, the calculated PSN edges are retained only if their I is greater than a defined cut-off value (I_{min}). ii) The calculation of the communication paths using information from both the PSN itself and

correlated motions (based on Linear Mutual Information (LMI) in this study). The contribution of the MD simulation to the approach comes in both the aforementioned steps, i.e. i) directly in the calculation of the PSN, where it is required to define a minimum number of frames of the MD ensemble in which each edge is observed; ii) in the path calculation, where the paths of communication are calculated for each MD frame using the I_{min} value obtained at the previous step. Moreover, each path is accepted only if at least one residue belonging to the path features correlated motions with one of the path terminal nodes higher than a defined threshold. Correlated motions are calculated from the MD ensemble using a specific correlation metrics, which is LMI in our case (see below). It becomes thus crucial to assess the dependency of the calculated paths from the selected I_{min} values and the values used as correlation cutoffs for the metrics employed to describe coupled motions in the MD ensemble.

The PSN for the different replicates and the evolution of the largest clusters as a function of different I_{min} values.

The PSN in the MD ensemble is the first component needed for our study. We first encounter the issue to define the minimum number of frames at which a contact (edge) between two residues is found at a distance lower than 0.45 nm. In other studies, included the ones by the original developers of the method, a value of at least 30 or 50% of frames is generally employed at this stage. We have thus decided to perform the PSN calculation using both these values and compare them.. In particular, we calculated several PSNs for each replicates of both arPTE-wt and arPTE-8M, varying the I_{min} cutoff from 0.2 to 40.0 by steps of 0.2. This provides a first analysis of the evolution of the largest cluster (cluster 1) size as a function of the I_{min} values (Figure 5.5).

Previous studies on crystal structures from PDB [34] pointed out that in most cases the size of the largest cluster undergoes an abrupt transition from higher to lower values. The midpoint of this transition (named I_{crit}) has been usually considered as the I_{min} value to be employed for PSN studies. As expected, the size of cluster 1 is larger at each I_{min} value when the 30% edge-cutoff is employed with respect to the usage of a

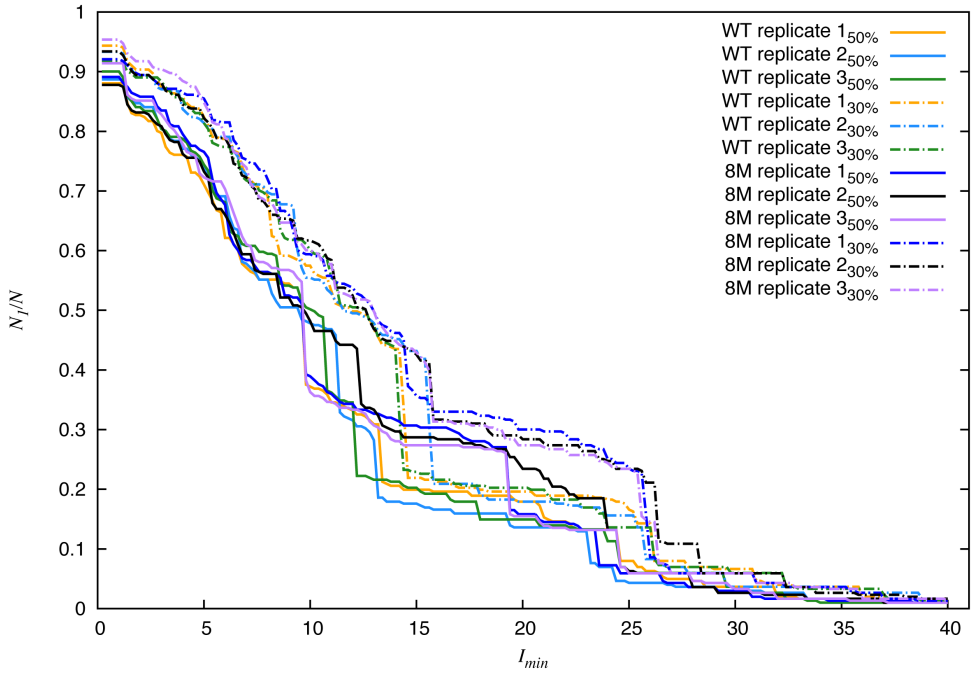

Fig. 5.5 Evolution of the size of the largest cluster as a function of the I_{min} values in the three replicates. In the legend on the right, the subscript values indicate the minimum fraction of frames in which a PSN edge has to be found. The plots here show that the choice of the 30 or 50% cut-off for the minimum fraction of frames in which a PSN edge has to be found has a limited impact on the calculation of the I_{min} value.

50% cutoff, due to a higher number of contacts retained within each PSN in the MD ensemble. The overall profiles are nonetheless quite similar at 30% and 50%, meaning that the choice of this cut-off value has minor influence, at least in this case, for the calculation of the I_{crit} value. Moreover, profiles of the I_{min} as a function of cluster 1 size are overall similar among the replicates of arPTE-wt and arPTE-8M, demonstrating the robustness of the PSN approach even when proteins with different sequences are compared. In Figure 5.5, it is possible to observe a first and steepest transition (I_{crit}) at an I_{min} of approximately 7 or 8 for the profiles calculated with stability cut-off of 50 and 30%, respectively. Nonetheless, transitions of minor entity, especially at low I_{min} values, can be identified. At higher values, especially from I_{min} equal to or larger than 10, the PSN becomes highly disconnected, with a huge number of very small and probably meaningless clusters. At these values it is reminiscence of a random graph, as previously observed in single PDB structures [34, 207]. Given the complex behavior of the evolution of cluster 1 size as a function of I_{min}, we decided to employ not a unique I_{min} value, but I_{min} values of 3.0, 4.0, 5.0 and 6.0 for further analyses.

Correlated motions are consistently described by Linear Mutual Information in the different replicates of arPTE.

To identify correlated motions in the simulations, we employed the Linear Mutual Information (LMI) [127]. It is a necessary step for the PSN-MD approach, which relies on the integration of the PSN with metrics that described correlated motions in proteins [89, 156]. LMI coupled to PSN have been applied for several different biological targets [156] and was thus a natural choice for our study. Moreover, LMI has the advantage of not depending on the relative orientation of the fluctuations with respect to the widely used Pearson cross-correlation. This characteristic of the LMI makes then possible to identify correlated motions unregard of the difference between their relative orientation in space [127]. An average LMI matrix was calculated for each replicate using time-windows of 2 ns. As shown in Figure 5.6, the average LMI matrices are very similar in the different replicates of arPTE. To quantitatively estimate the simi-

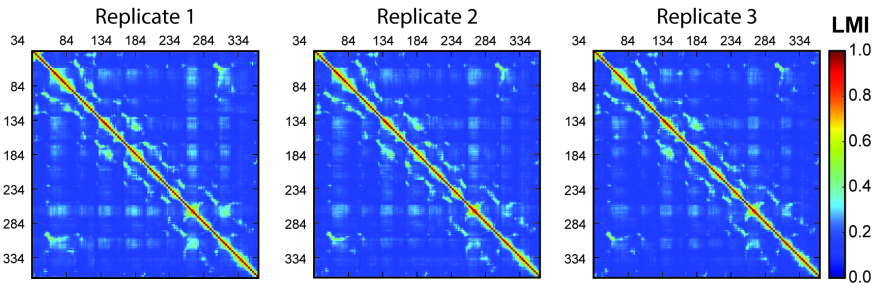

Fig. 5.6 Average LMI matrices calculated for each of the three replicates of arPTE-WT, using time-windows of 2 ns for the averaging. From these figures it is evident how the three calculated LMI matrices are similar among the different replicates of the wild-type.

larity between the LMI matrices of the different replicates, we calculated the pairwise Frobenius norm between them (see Materials and Methods), which ranges from 6.973 to 8.017. As a baseline, we also calculated the Frobenius norm from the average LMI matrices from the two halves of each independent trajectory, collecting values between 7.356 and 9.506. The overall good agreement between the inter-replicate and intra-replicate comparisons suggests that the three replicates feature overall similar patterns of correlated motions. The average LMI here calculated can thus be employed to estimate the paths of communication by the PSN-LMI approach ([14]).

Dependence of the paths of communications from the LMI cutoff in the PSN.

As detailed above, the calculation of the communication paths in the PSN-MD method relies on information from both the PSN and the metrics used to described coupled motions between residues in the MD ensemble. Both the calculation of the PSN and the correlated motions require the definition of a significance cutoff so that we can discard paths that are not relevant and can be ascribed to noise. Thus, to properly identify a threshold for LMI values, we calculated the communication paths as a function of LMI cutoff values for each of the three replicates, using LMI cutoffs ranging from 0.25 to 0.7 by steps of 0.05 (Figure 5.7). For these

calculations, the I_{min} value was varied from 3.0 to 7.0 in discrete steps of 1.0. The number of paths for each pair of mutation site/loop 4 or 7 residues, as well as the total number of paths were calculated for each replicate of arPTE-wt. As expected, the total number of paths decreases at increasing values of the LMI cutoff (Figure 5.7). The steepest transition is observed for LMI cutoffs of approximately 0.4 in almost all the cases. At high I_{min} values (as for example 6 or 7), the number of paths is less influenced by the LMI cut-off value. This may be mostly related to the fact that the usage of a high I_{min} cut-off filters out most of the paths by itself. In light of these results, we selected for further analysis a LMI cut-off of 0.4.

Communication paths in the arPTE-WT and arPTE-8M replicates.

Once the value of the LMI cutoff has been set at a value of 0.4, we then focused our attention on the dependence of the paths from the I_{min} value. We thus calculated the paths at four I_{min} values (3.0, 4.0, 5.0, 6.0), discarding all the paths with a probability of occurrence lower than 15% to discard poorly representative paths. It should be noted that the X-ray structure used as reference for the wild-type variant (arPTE-WT) actually presented a mutation (N265D) also present in the 4M variant only. The 8M variant, instead, did not feature the mutation. This means that residue 265 was 265Asp in arPTE-WT, and 265Gln in arPTE-8M.

Most mutation sites are involved in long-range communication to loop 7 and 4 in arPTE-WT, with few exceptions. Among the ten mutations (I260T, N265D, G273S, T274N, K185R, D208G, Q206P, G60A, A80V, R118Q), which are present in 8M, 4M arPTE variants or in both, two of the 8M mutation sites are glycine residues in the wild-type (Gly273, Gly60) that are not predicted to be involved in any path since the PSN does not account for glycines. Of the remaining mutation sites (Ile260, Asp265, Thr274, Lys185, Asp208, Gln206, Ala80, Arg118), most of them (Ile260, Asp265, Thr274, Lys186, Asp208, Ala80) are involved in communication paths with the residues in the surroundings of loop 4 and loop 7, which are the ones differing in orientation in the E_{open} and E_{closed}

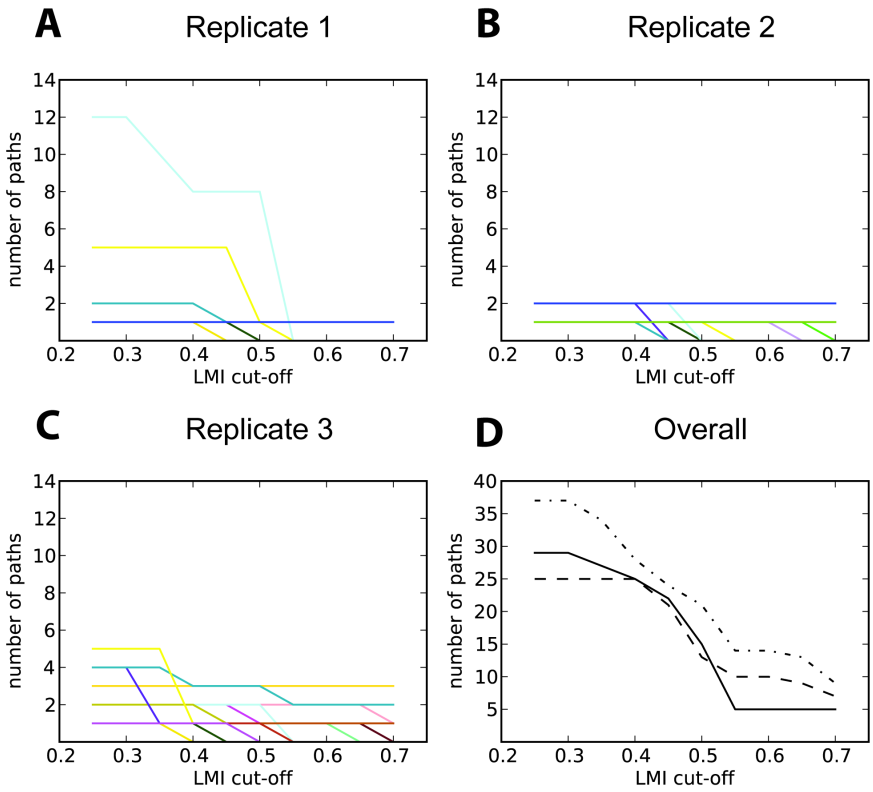

Fig. 5.7 Dependence of the number of communication paths from the LMI cutoff values. A-C) The shortest communication paths between each mutation site and each of the residues in loop 4 and 7 are showed with different colors. The number of paths are shown for each replicate as a function of the values of the LMI cutoff. D) The total number of paths identified between mutation sites and residues in loop 4 and 7 is shown for each replicate as a function of the values of the LMI cutoff. The values allowed to choose for an LMI cut-off of 0.4.

states (Figure 5.1). Gln206 and Arg118 instead are not involved in any paths. In this case, a more local effect due to an alteration of the native intramolecular interactions cannot be ruled out, since these residues are located close to the loop 4 region and Phe132. From a structural point of view, we have divided the mutation sites in three main classes: those belonging to loop 7 (Ile260, Asp265, Thr274, Gly273), those belonging to the loop 4 region (Lys185, Asp208, Gln206) and other mutation sites (Gly60, Ala80, Arg118). The latter group is composed of three mutations, which are located in close proximity each other. This region spans loop 1 (Gly60), helix 1 and helix 2 (Ala80 and Arg118).

L7 mutation sites communicate exclusively with other L7 residues. As previously introduced, three mutation sites (Ile260, Thr274, Asn265) among those present on loop 7 are were found to be involved in communication paths, even though paths from Thr274 are present only at I_{min} 3 (Figure 5.1-A). Interestingly, the L7 mutation sites are connected by paths exclusively to other residues of loop 7, i.e. in the same region of the protein where they are located. Indeed, at I_{min} 3 these three mutation sites are connected to each of the L7 residues involved in the conformational transition in at least two of the simulation replicates, through more than one path in most cases for each pair of residues. These paths feature intermediate nodes that are located on loop 7 and on a short helical secondary structure element which is embedded in the long loop 8 (residues 313-320). By increasing the I_{min} cut-off for paths identification, the paths characterized by weaker interactions, such as those involving Thr274, are discarded. Nonetheless, some of the pathways involving residues Ile260 and Asp265 are retained up to I_{min} 6. The localization of the identified paths enforces the notion that the aforementioned mutations act on the protein dynamics on a relatively local scale.

Mutation sites on the L4 region are involved in both local and very long-range paths. Among the L4 mutation sites, residue Gln206 is not involved in any path. Instead, both Lys185 and Asp208 feature communication paths. In particular, Lys185 features a very high tendency to communicate, with paths being present when using the strict I_{min} 6 cut-off. Paths connecting mutations of the L4 region with residues in the L4

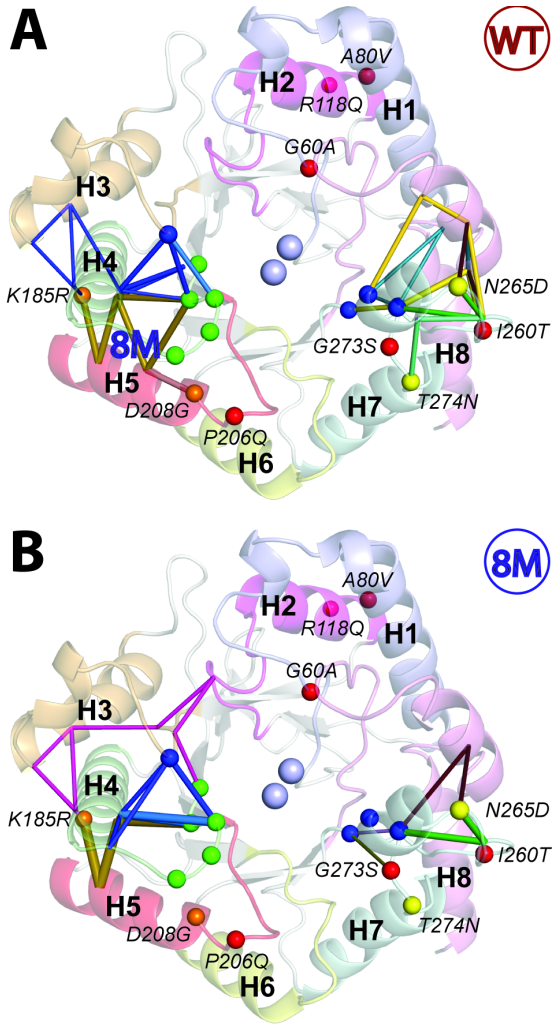

Fig. 5.8 identified interaction paths connecting mutation sites on L7 with L7 residues, and mutation sites on L4 with residues of the L4 region. These figures show the interaction paths for two representative replicates of arPTE-WT and arPTE-8M. The paths connecting two given residues are shown as chains of cylinders connecting the terminal nodes as well as the intermediate ones; cylinder width is proportional to the frequency with which a given path was found. All the identified paths connecting two given residues have the same color, which is different for every residue pair; notice that some cylinders may be superimposed, because certain residue pairs share part of their paths. In these figures, the Cαs of mutation sites present in the arPTE-8M mutant only are shown as red spheres, those present in the arPTE-4M mutant only are shown as yellow spheres, and mutations present in both are show as orange spheres. Residues involved in the conformational transition are colored in blue, while those of loop 4 (171-175) are shown in green. Helices and their respective loops are colored in order to be able to distinguish them in the structure.

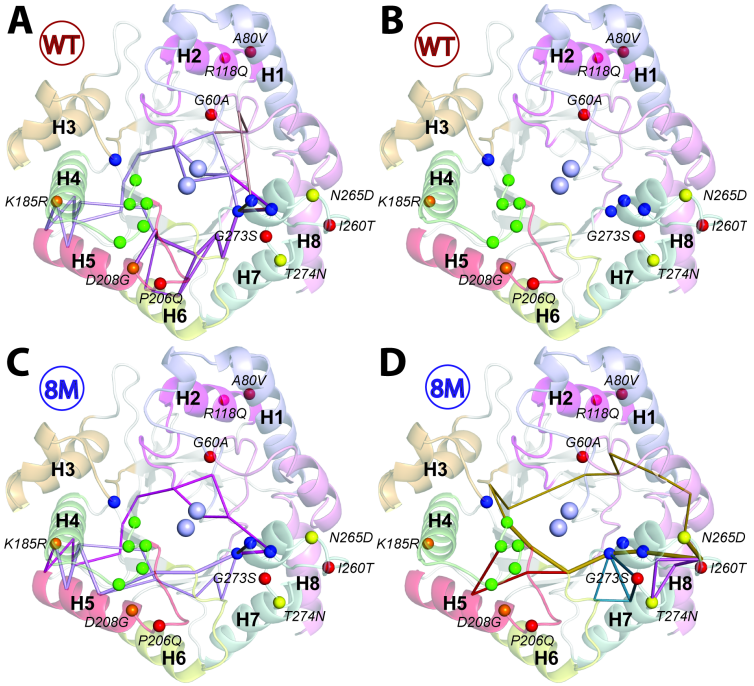

Fig. 5.9 identified interaction paths connecting mutation sites on L4 with L7 residues (panels A and C) and connecting mutation sites on L7 with L4 residues (panels B and D) in all the replicates of arPTE-WT and arPTE-8M. Notice that no paths are present in panel B, because the arPTE-WT simulations didn't feature any for the specific class. See legend of Figure P1 for further details about coloring and representation.

region itself involve the alpha helix where Lys185 is located and adjacent secondary structure elements, most prominently the helix 5 (Figure 5.8-A). Interestingly, L4 mutation sites are also involved in communication paths with residues of L7, through very long paths which cross extensive parts of the protein structure (Figure 5.9). These paths follow two different routes: those starting from Lys185 cross the enzyme structure through the alpha-beta loops at the opposite side respect to the active site and then pass through the beta-alpha loops 2, 1, 8 and 7, landing on Arg254. Paths starting from Asp208 towards L7 follow a more direct route, passing through L5, L6 and finally reaching L7.

Ala80, in the H1/H2 region features robust, long-range paths with

both L7 and L4. Three mutation sites are located in the helix 1/helix 2 region: Ala80 and Arg118 are spatially very close with one another, and are located, respectively, on helix 1 and 2. Instead, site Gly60 is located on loop 1, and is thus closer to the L7 region. Of these residues, Ala80 alone features paths of communication with the rest of the protein (Figure 5.10-A). The path connecting Ala80 with L7 involves L8, L1 and close-by secondary structure elements, such as helix 8 and helix 1. Paths connecting Ala80 to L4 residues pass instead through the opposite side of the protein and involve instead part of the beta-strand of PTE's tim-barrell, loops 2, 1 as well as helix 2 (Figure 5.10-B). A less direct route also has been identified in replicate 2 only, crossing the central part of the tim-barrell, involving the Zn^{2+} atoms and reaching the L7 loop residue Arg254. From that point, the communication path reaches Ala80 through the previously described L7 mutations-Ala80 path.

The arPTE-8M mutant features overall similar communication paths as the wild-type, although important differences arise. The mutant residues in arPTE-8M feature, for a good part, similar communication patterns as in the wild-type. However, important differences can be identified in the calculated paths. Among the L7 mutations, the communication propensity of residues Asn265 and Thr274 are mostly unchanged. Nonetheless, the introduction of the I260T mutation makes this residue less prone to be found in paths, since, unlikely as happens in the wild-type simulations, communication paths involving it are found only in replicate 3 and at I_{min} 3. Moreover, the G273S substitution allows for residue 273 to be involved in paths, even if at I_{min} values of 3 or 4. Considering the mutations in the L4 region, Asp208 is replaced by Gly, which cannot be involved paths. The Q206P mutation does not increase the residues communication propensity, i.e. the residue feature no paths both in the wild-type and in the mutant variant. When the H1/H2 mutations are considered, it is possible to notice an increased communication propensity with the inserted mutations. Indeed, the G60A mutation allows for residue 60 to be involved in communication paths, which subsist even when the high I_{min} 6 value is used. Moreover, the mutation of residue 118 from Arg to Gln greatly increases its communication propensity, as this residue is

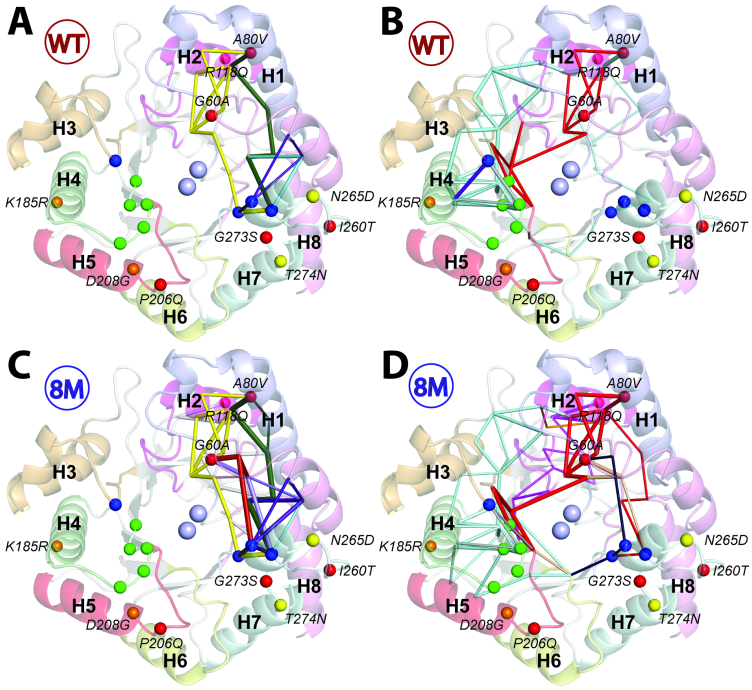

Fig. 5.10 identified interaction paths connecting mutation sites on H1/H2/L1 with L7 residues (panels A and C) and L4 residues (panels B and D) in all the replicates of arPTE-WT and arPTE-8M. See legend of Figure P1 for further details about coloring and representation.

involved in paths at both I_{min} 3 and 4, which haven't been found in the wild-type simulations.

L7 mutation sites communicate both with L7 and L4 residues in arPTE-8M. As previously described, the L7 loop mutations communicate exclusively with L7 in arPTE-WT (Figures 5.8-A and 5.9-B). The observed behavior is slightly different in arPTE-8M. More in detail, paths communicating from mutation sites on L7 to L7 itself are diminished in number in the arPTE-8M simulations (Figures 5.8-A and 5.9-B). Moreover, the arPTE-8M simulations feature a higher number of relatively weak paths, which are found only at I_{min} 3 or 4, with the residues of loop 4. These paths follow similar routes as the paths from the L4 mutation sites to L7 residues, passing through L4, L5 and helix 5, and are exclusive of the 8M simulations alone (Figures 5.9-B and 5.9-D).

The D208G substitution affects the ability of mutation sites in the L4 region to communicate. In general, mutation sites in the L4 region communicate less in arPTE-8M respect to the wild-type, mostly because of the D208G mutation. The absence of residue 208's side-chain affects all the paths in which residue 208 is involved, thus diminishing the total number of paths that the mutations in this region are able to connect to other parts of the protein. This effect affects both local (i.e. connecting the mutations of the L4 region with L4 itself and Phe132 - see Figure 5.8-B) or very long-range paths (Figure 5.9-C). Nonetheless, the conservative mutation K185R does not modify greatly the communication propensity of residue 185, which communicates with the L7 residues as well. However, the disappearance of the paths mediated by D208 in arPTE-WT changes the communication routes with L7. Indeed, the pathways cross a more central region of the protein, traversing the tim-barrel's beta-strands, loops 5 and 6 and also the Zn^{2+} atoms and surrounding regions (Figure 5.9-C).

Mutated sites at the H1/H2 and L1 communicate visibly more in arPTE-8M. Of the three mutation sites belonging to the helix 1-helix 2 region in arPTE-WT (Gly60, Ala80, Arg118), only one (Ala80) is able to communicate with other residues of the enzymes (Figures 5.10-A and 5.10-B). Instead, in the 8M mutants, all three mutated sites (Ala60, Val80,

Gln118) communicate with both the L4 and L7 residues (Figures 5.10-C and 5.10-D). In fact the number of communication paths which connect both L4 and L7 to the three H1/H2 region residues is increased in the mutant respect to the wild-type variant. Nonetheless, the main communication routes to residues Val80 and Gln118, which are very close one another, remain similar to the ones found in the wild-type. As far as Ala60 is considered, the paths connecting it to loop 7 pass through the close-by loop 8. Instead, those reaching loop 4 pass either through the central region of the tim-barrell or connect to L7 first, then reaching Ala60 through L8. The latter paths, however, are found only at I_{min} 3 or 4.

Structural communication at the Glycine mutation sites

As previously pointed out, the PSN/MD approach was originally conceived so that it cannot account for glycine residues, since it is based on side-chain contacts. Moreover, glycine residues have been often employed as mutation to interrupt long-range communication thanks to their intrinsic flexibility and the absence of side-chain mediated interactions. To verify if structural communication routes mediated by Gly60 and Gly273 residues in arPTE are in act, we decided to apply a network analysis to the average LMI matrices directly (see Materials and Methods). In this application, the edges have weights equal to the correlation value embedded in the LMI matrix. The shortest communication paths have been calculated between Gly60/Gly273 and the residues related to the E_{open}/E_{closed} conformational transition. Figure 5.11 shows the paths for the replicates of arPTE-WT, which are overall quite consistent. Notably, the two Gly residues mediate communication paths toward the residues on the side of both loop L7 and on the side of loop L4. While some variability is indeed expected, because the three replicas sample not completely superimposable regions of the conformational space, paths mostly pass through the same regions of the protein in the three replicas.

Gly60 (mutation G60A) is located on the loop L1. This residue can mediate long-range communication to loop L7 through loop L8 (residues 299-326). Two main routes are instead available to connect Gly60 to loop L4 and the surrounding region. Indeed, Gly60 can communicate with

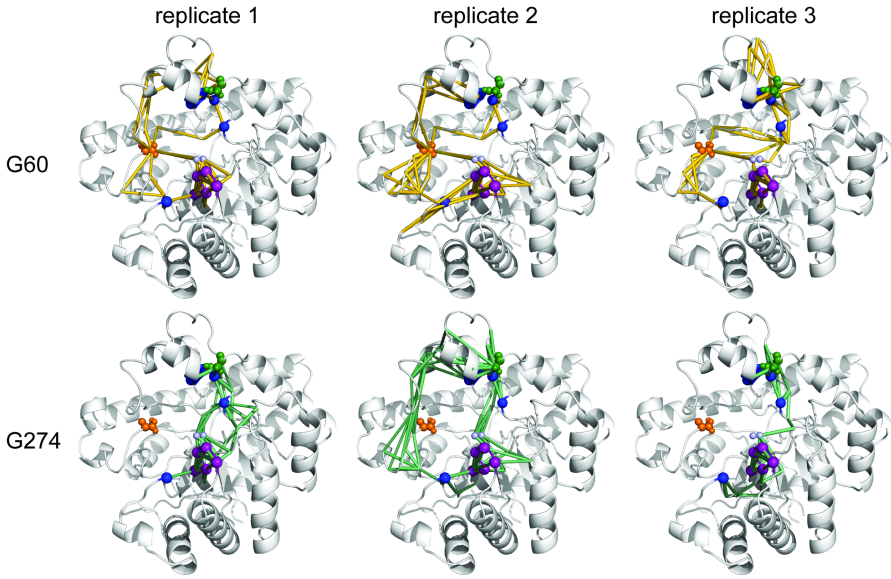

Fig. 5.11 Paths identified in the filtered LMI matrices for residues Gly60 and Gly273. Communication paths exist in the LMI matrix connecting these Glycine residues to those involved in the E_{open}/E_{closed} transition. Paths connecting residues Gly60 and Gly273 are shown, respectively, as orange and green cylinders, while residues themselves are shown as spheres. The C$\alpha$of the residues involved in the conformational transition are show as blue spheres, with those of loop 4 shown as purple spheres. Zn2+ atoms are shown as light blue spheres. Several paths have been identified connecting the Gly mutation sites with the rest of the protein.

loop L3 - where Phe132 is located - and loop L4 through residues in loop 2. An alternative route involves the the Zn^{2+} ions and then loop L5. The metal ions are also directly included in the PSN/MD approach as well„. and few paths comprising the active-site's metal ions were also identified using that approach, although at low I_{min} values. Nevertheless, we fail in identifying any pathways passing through them with the other mutation sites. It is hard to conclude if this is due to a specific Gly-mediated communication or if it is just related to the different approaches employed, which relies on a diverse nature of the underlying network (i.e. PSN vs LMI) Gly273 (mutation G273S) is located on loop L7, next in sequence respect to residue Phe272. Mutating it to any other residue would by itself modify the local dynamical patterns, by restricting the conformational freedom at this position and influencing the whole loop. This residue is characterized by uncorrelated motions with respect to the the surrounding residues in the wild type, and thus no local paths are identified between Gly273 and the loop L7 surroundings. Not surprisingly, paths do exist between this residue and the others of the L7 region. Nonetheless, we can identify paths of communication between Gly273 and loop L4 through loops L6 and 5 and the Zn^{2+} atoms. In summary, the existence of communication paths connecting the Gly60 and Gly274 mutation sites to loops L4/L7 surroundings highlights a potential effect on protein native dynamics induced by mutations at these sites over long distances.

Local effects of mutations that are not involved in long-range paths.

The analysis of communication paths using the PSN method highlighted that the residue at position 206 (mutation Q206P) doesn't feature any communication paths with the rest of the structure in the wild-type. Nonetheless, due to the peculiar position of the mutation, we hypothesized that the substitution might have a local role in modifying the dynamics of the L4 region. Indeed, residue 206 is located on L5, which is close enough to L4 to be able to interact with it. We have thus performed the analysis of non-bonded interactions for this residue in the crystal structures

of WT and 8M arPTE, as well as on the simulation trajectories. In the crystal structure of the wild-type variant (PDB ID 2R1N), residue 206 is involved in a network of hydrogen bonds which involve residues from L5 as well as L4. Indeed, an hydrogen bond is present between Gln206 Nε-H atom and the backbone carbonyl group of Thr173 (L4). Residue 206 is also engaged in other hydrogen bonds through the Nε-H and the N-H backbone atoms, respectively with the side-chain terminal hydroxyl group and backbone carbonyl of Ser203 (Figure 5.12-A).

Interestingly, the Q206P mutation completely depletes the ability of residue to 206 be involved in such hydrogen bonds, so that, the interactions both between L5 residues and with residues of L4 are disrupted (Figure 5.12-B). A concurring factor is also the D208G mutation, since residue Asp208 interacts with the L4 residue Gly174 through an hydrogen bond. The analysis of the simulation trajectories of the wild-type (Figure 5.12-C) shows, however, that the Gln206-Thr173 hydrogen-bond is not present in most of the simulation frames (<5% of persistence), and that residue Asp208 does not strongly interact with residue 173 as well (Figure 5.12-D). More in general, it appears evident the the Q206P mutation changes significantly the interaction and local dynamics of loop 5, but it is less clear than in the crystal structures how these differences propagate to loop 4, since the differences in the L4-L5 interaction in the wild-type and mutant variant are not as striking as the crystal structures would suggest. It should be noted however that the considered region is quite well solvent exposed, meaning that the solvent model used might influence the obtained results. In order to confirm said results, It would be advisable carrying out a MD simulation with the most appropriate solvent model for the CHARMM22* force-field.

Summing up, the differences in the local interactions between arPTE-WT and arPTE-8M in the simulations are less pronounced than in the crystal structure: the changes in the local interactions surrounding residue 206 do have an effect on the interactions with loop 4, although to a greatly reduced extent than what appears from the crystal structures. We also turned our attention to the R118Q mutation. Indeed, residue 118 is not involved in any communication paths in the wild-type, while it is in

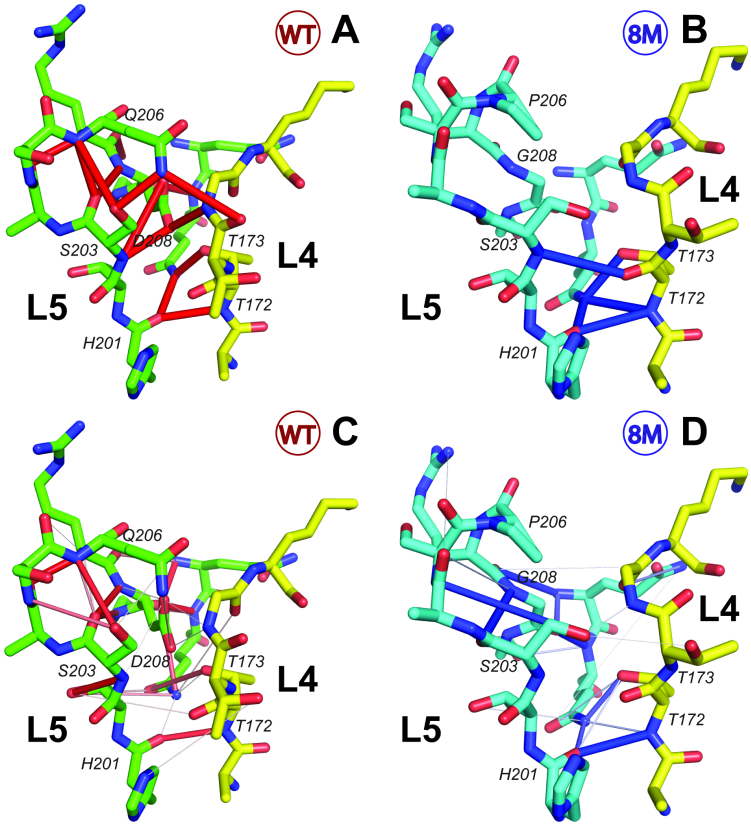

Fig. 5.12 Hydrogen bonds identified in the arPTE-WT and arPTE-8M variants in the L4-L5 region, in two representative replicates. Panels show hydrogen bonds identified in the wild-type and 8M PTE variants as respectively blue or red cylinders, connecting donor and acceptor atoms. Cylinder width and color saturation are proportional to their persistence in the ensembles. Hydrogen atoms are not shown for clarity. Panel A and B show hydrogen bonds identified in the crystal structures (PDB IDs 2R1N and 3A3W) of, respectively, arPTE-WT and arPTE-8M, while the lower panels C and D show the hydrogen bonds identified from two representative replicates among those calculated. L5 is colored in green or cyan for respectively the WT and 8M variant, while L4 is colored in yellow for both. While the interaction identified between the L4 and L5 residues are strikingly different when considering the wild-type and 8M variant, the differences identified in the simulations are less relevant.

the arPTE-8M mutant. It has to be noted that in all the communication paths having residue 118 as terminal residue in arPTE-8M, residue Gln118 is connected first to residue Arg41 or Val40. In arPTE-WT, in which residue 118 is an Arg as well, the electrostatic repulsion between the two positively-charge side-chains is unlikely to allow interaction between the respective side-chains. Besides, Arg118 is involved in a salt-bridge interaction (persistence of 47.9 to 77.7 in the three replicates) with Glu115, which is on the opposite side respect to where Arg41 is located.

5.4 Discussion

It has been proposed that the reaction rate of arPTE is remarkably influenced by the equilibrium between an “open” and “closed” conformations, that differ by the orientation of few residues in the proximity of the active site. Mutations at distal sites with respect to the catalytic region (as the ones observed in the so-called 4M and 8M mutants) affect both the relative population of the "open" and "closed" states [110] and the kinetic parameters of the enzyme [110]. Overall, these data suggest that remote mutations can affect the turnover rate of arPTE by altering its conformational landscape. The long-range communication between the mutation sites and the residues in the proximity of the active site can thus play a role. In this context, we here used arPTE as a model system to test if methods inspired by graph analysis and coupled to molecular dynamics (MD) simulations are able to detect paths of long-range communications between the known mutation sites and the residues in proximity of loops 4 and 7, which are the hallmark of the conformational change between "open" and "closed" states. Indeed, the availability of experimental structures for the two conformations and the knowledge of their relationship with protein activity provides us with a valuable source of information to assess the capabilities and robustness of PSN approaches coupled to MD, as well as to provide atomistic details on the residues involved in the conformational change. At first, our results highlighted some of the inherent limitations of the PSN/MD method. It is encouraging for the application of this kind of methods to see that we can find a large agree-

ment for the structural location of the paths between each mutation sites and the residues in the proximity of the catalytic site, especially when the method is coupled to a new generation and experimentally-validated force field as the one employed here. Nevertheless, the three replicates feature some differences in the number of the total paths identified between each pair of sites and often different values of I_{min} cutoff are needed to get a similar number of paths in the different independent trajectories. Moreover, the identity of the intermediate residues in the paths is not always strictly conserved. While the simulations sample overall similar conformational spaces, these are not completely overlapping. The identified differences thus arise from the fact that the protein structure network analysis, due to its great sensitivity on the sampled conformational distributions, amplifies the differences that exist between the conformational spaces sampled by the different replicates. Even though the MD simulations were found to sample similar conformational spaces and close to the experimental structure, meaning that these networks are very sensitive to the conformational distribution. Moreover, the PSN/MD approach relies on the choice of several cut-off values, especially for the step in which the metrics for correlated motions are taken into account (LMI in our application) and for the I_{min} value of the PSN itself. Their selection is not always very clear-cut, as also pointed out very recently in another study. It is thus necessary, approaching to these studies, to test different parameter values and carefully verifying the reproducibility and robustness of the results, as we here did for the PSN I_{min} and the correlation values from LMI matrices. Our study then shows that several communication paths can be identified from the distal sites to the key residues in the surrounding of the catalytic site, in agreement with the fact that the long-range effects in protein structure during dynamics can be propagated using multiple paths that have often nodes in common. Two of the mutation sites are glycines (Gly60 and 273) and the PSN method, as it is conceived, does not allow to account for glycine residues in the protein structure graph. Indeed, glycines are a special class of amino acids being deprived by a side chain and they are not expected to be involved in the propagation of the communication in terms of cascade of steric clashes

between side-chains. Of course, this might not be the only source of structural communication, since structural long-range effects can be due to other factors. We here decide to focus on the information that the PSN approaches can provide so the usage of other kind of methods is out of the original scope but they have to be considered in the interpretation of the results and for future studies. Nevertheless, Gly60 and 273 are located, respectively, on loop 1 and 7; since close-by residues (as those on loop 7 or Ala80, respectively) are involved in communication paths, we used graph analysis on the LMI matrices in order to search paths of communication which did not strictly depend on side-chain contacts. Indeed, the analysis identified several communication pathways with residues important in the E_{open}/E_{closed} transition. This suggests that the PSN-PATH analysis alone would benefit by being complemented with techniques that do not strictly rely on side-chain contacts and that take into account other types of structural communication.

In our analyses, Ala80 turns out to be one of the main distal sites for long-range communication and these paths are still found even using very strict cut-offs such as high I_{min} values. Nonetheless, other long-range communication pathways were identified, involving residues closer in structure. In particular, the mutation sites on L7 were found to be connected only with L7 residues implicated in the conformational transition in arPTE-WT. We also show that two residues (Gln206 and Arg118) belonging to the mutation sites of 8M and 4M variants are not engaged in any paths at any used I_{min} values. Since Gln206 is spatially close to the loop 4, the mutations in this position may exert a local effect. Indeed Gln206 is at around 8 Å of distance from loop 4 in the crystallographic structure if Cαalpha-Cαalpha distances are considered. Thus, we hypothesized that its mutation to proline causes the loss and/or perturbation of local intramolecular interactions and dynamics. To investigate whether this was the case we analyzed the network of intramolecular interactions mediated by these two residues in both the x-ray structures and in the MD ensembles of both wt and 8M mutant arPTE. The analysis of the interactions in the x-ray structures pointed out key differences in the hydrogen-bonds networks of the area sorrounding residue 206 between

the two variants, also affecting the hydrogen bonding capability of these residues with residues of L4 itself. However, the analysis of the MD trajectories showed that, while some differences are present, they are not as relevant as was evidenced by the X-ray structures alone, meaning that the interactions between the L5 Gln206 and L4 are not strikingly different in the wild-type and mutant variant. This suggests that the Q206P mutation itself might have a minor role in influencing the difference in the conformational distribution observed between the arPTE-WT and arPTE-8M variants. The situation is different when the R118Q mutations is considered. In this case, replacing one residue with one with a smaller side-chain results in increasing its predisposition to be engaged in communication paths. In this case, a change in its chemical-physical properties, by replacing a positively charged Arg residue with polar Gln, allowed its side-chain to be involved in a different set of interactions which resulted in the activation of a communication pathway.

We also investigated also the path distribution in the simulations of the 8M mutant by using the same PSN/MD approach. Indeed, our analysis highlighted some key differences between the wild-type and the mutated variant. arPTE-8M features a lower number of paths connecting mutation sites with close residues involved in the E_{open}/E_{closed} transition. This happens both in the L7 and L4, where key mutations (such as D208G) inhibit the formation of communication paths. Instead, a higher number of very long-range paths, that cross different regions of the proteins, have been identified in the 8M variant. This happens mostly because key substitutions in the L1/H1/H2 region, which is far form both L4 and L7, allow to sensibly increase the communication with both the L7 and L4 region. Two of these mutations (G60A, A80V) represent an increase in the size of their side-chain, allowing them to achieve better contact with the surrounding protein. This is especially relevant for residue 60, as the presence of a side-chain, albeit as small as that of alanine, results in communication pathways present even at the high I_{min} 6 filtering. In one case (R118Q), the substitution of a charged for a polar residue allows for a different set of interactions to take place, as detailed before, thus activating a communication route which is not present in the

wild-type.

5.5 Conclusions

Summing up, we have been able to identify multiple communication routes between mutation sites and residues involved in E_{closed}/E_{open} transition, from most of the identified mutation sites. By comparing the dynamical properties of the arPTE-WT and arPTE-8M simulations we were able to consistently identify quantitative differences between the two systems, especially in the number of very long range communication paths between the L7 residues and the helix 1/helix 2/loop 1 region. Indeed, a different set of interactions resulting in a cascade of effects acting on the active site might be able to influence the conformational equilibrium between the "open" and "closed" forms. Of the considered mutation sites, residue 260 (mutation Q260P) has been found to be not involved in any paths in the calculated paths. While the substitution is likely to influence the very local dynamics, its effects on L5 itself are less evident in our simulation than in the crystal structures, suggesting that this mutation may have a minor role in influencing the open/closed transition.

Chapter 6

Conclusions

In the present work, we have investigated the function-structure relationships in three different metalloenzymes: the Mo-binding formate dehydrogenase (FDH), the [NiFe] hydrogenase and Zn-binding phosphotriesterase. These enzymes have been chosen both for the great relevance of the catalyzed reaction in a green economy context, as well as for the ability to inspire related biomimetic chemistry and their involvement in bioremediation.

For the two metalloenzymes involved in redox reactions, density functional theory was used. In FDH, we investigated potential intimate mechanisms associated with the rate-determining step of the $HCOO^- \rightleftharpoons CO_2 + H^+ + 2\,e^-$ reaction, by testing several reaction mechanisms in which different atoms of the active site could act as hydrogen acceptors. An accurate model of the MGD ligand has been employed, in order to fulfill the requirements for the correct treatment of the cofactor stereoelectronic features. Among all investigated mechanisms, those featuring an initial hydrogen migration as hydride from the substrate to the metal ion have turned out to be the lowest in energy. Notably, the conserved Se-Cys retains an important role because it is involved in shuttling protons from and to the metal cofactor. The observation that part of the electron density arising from $HCO_2{}^-$ oxidation is localized on the pyrazine moiety of the cofactor, as well as localization of HOMO/SOMO on the same region, confirms that employing an extended model of the cofactor is mandatory

for describing properly the chemistry of FDH and MPT binding enzymes in general.

We then turned our attention to another promising target of [NiFe]-hydrogenase. We have used DFT to study the H_2 binding and activation properties of models of the dinuclear metal center. Our findings highlight that the spin state and the coordination geometry of the Ni ion are two crucial factors that tune the energetics and regiochemistry of H_2 binding to the NiFe metal cluster, thus suggesting that the hydrogenase is in an entatic state. In particular, H_2 binding is energetically favored only when the NiFe metal cluster is in a low spin state and the Ni cysteine ligands are arranged in a distorted seesaw coordination geometry, as observed in the crystallographic structures. In addition, when the Ni coordination sphere has such peculiar stereoelectronic features, H_2 binding takes place exclusively at the Ni center. Two low activation energy pathways are possible after H_2 binding to the Ni(II) ion: direct heterolytic cleavage of H_2mediated by the sulfur atom of one of the terminally coordinated cysteine residues or oxidative addition of H2 followed by proton transfer to the sulfur atom of one of the terminally coordinated cysteines. The latter reaction pathway is almost barrierless, indicating that H_2 cleavage can occur very quickly. From an electronic structure perspective, our results complement previous observations which highlighted how an increased driving force for H_2 addition is linked to the distortion of square planar Ni complexes, essentially because such a distortion leads to a decrease in the energy of the LUMO, which enhances the hydride acceptor ability.

Phosphotriesterase, instead, is implicated in the hydrolysis of organophosphates and related compounds. As previously detailed, it has been proposed that the reaction rate of phosphotriesters in arPTE is heavily influenced by the equilibrium between an "open" and "closed" conformations, that differ by the orientation of few residues in the proximity of the active site. In order to gain insight on the structure-function relationships for this critical step, we have studied the structural communication routes between the residues of the active site implicated in the conformational transition and the mutation sites both in the wild type and in the 8M mutant variant. Our analysis has shown that several paths of communi-

cation between the mutation sites and active sites residues do exist, both in the wild-type arPTE-WT and in the mutant arPTE-8M. Nonetheless, the analysis allowed to point out peculiar communication routes as well as to outline key differences between them, and to point out which mutation sites were the most and least implicated in structural communication paths. When comparing the wild-type and mutant variants, we found that communication paths in arPTE-8M are overall similar to those of arPTE-WT, although with few important differences. In particular, arPTE-8M features a lower number of paths connecting mutation sites with residues close-by involved in the E_{open}/E_{closed} transition. This happens because key mutations inhibit the formation of local communication paths. Instead, a higher number of very long-range paths, that cross different regions of the proteins, have been identified in the 8M variant, mostly because key substitutions in the L1/H1/H2 region allow to sensibly increase the communication between the L4 and L7 regions. Indeed, a different set of interactions resulting in a cascade of effects acting on the active site might be able to influence the conformational equilibrium between the "open" and "closed" forms.

In the present work, we have used computational methods to investigate structure-function relationships in three chosen metalloenzymes related to the energetic and environmental challenges of the green economy. We were able to gain insight on the controlling step of the catalytic reaction for the three enzymes. For FDH and [NiFe]-hydrogenases we proposed novel reaction mechanisms which take into account the peculiar structural features of the respective active sites, and to relate the found mechanisms with pre-existing knowledge about the enzymes and related inorganic complexes. Of course, further work is necessary to deepen our knowledge on these systems. For instance, at least the whole catalytic cycle should be evaluated in order to further clarify the structural and stereoelectronic properties of all the intermediates. The accuracy of calculations could be increased by either including larger models of the active sites, especially in the case of FDH, or using hybrid approaches such as QM/MM. In phosphotriesterase, we were able to gain insight on the structural communication routes in both the wild-type and 8M mutant.

Indeed, our results show that even remote mutations are able to modify the structural comunication patterns of the enzyme and this likely to influence the different equilibrium between the E_{open} and E_{closed} states which is found in the two variants. We were also able to point out the mutations which communicate most with the L7 and L4 residues, and also mutations that are unlikely to affect the dynamics of active sites. However, our work does not straightforwardly proves that the identified mutations are responsible for the observed change in reaction rates and conformational landscapes. In order to confirm this, more experimental as well computational work needs to be carried out. Indeed, an interesting proposition would be to use a biased sampling method, such as metadynamics, in order to characterize the different conformational landscape of the arPTE-WT and arPTE-8M variants. A successive step would require to use the same procedure on mutants featuring subsets of the mutations present in 8M or 4M, aiming to determine whether the mutations identified as most communicating are in fact important in influencing the conformational equilibrium between E_{open} and E_{closed}.

References

[1] The enzyme nomenclature website, which is the authoritative website for EC by the nomenclature committee of the international union of biochemistry and molecular biology, can be found at http://www.chem.qmul.ac.uk/iubmb/enzyme/ .

[2] Abbas Abou Hamdan, Sebastien Dementin, Pierre-Pol Liebgott, Oscar Gutierrez-Sanz, Pierre Richaud, Antonio L. De Lacey, Marc Roussett, Patrick Bertrand, Laurent Cournac, and Christophe Leger. Understanding and tuning the catalytic bias of hydrogenase. *Journal of the American Chemical Society*, 134(20):8368–8371, May 2012. WOS:000304285700021.

[3] Neus Agell, Oriol Bachs, Nati Rocamora, and Priam Villalonga. Modulation of the ras/raf/MEK/ERK pathway by ca2+, and calmodulin. *Cellular Signalling*, 14(8):649–654, August 2002.

[4] R. AHLRICHS, M. BAR, M. HASER, H. HORN, and C. KOLMEL. Electronic-structure calculations on workstation computer - the program systema turbomole. *Chemical Physics Letters*, 162(3):165–169, October 1989. WOS:A1989AX76500001.

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

[6] Marta E. Alberto, Tiziana Marino, Nino Russo, Emilia Sicilia, and Marirosa Toscano. The performance of density functional based methods in the description of selected biological systems and processes. *Physical Chemistry Chemical Physics*, 14(43):14943–14953, 2012. WOS:000310005700027.

[7] A. ALBINATI, A. TOGNI, and LM VENANZI. Cationic molybdenum(iv)- and tungsten(iv)-iridium(iii) complexes with hy-

dride and eta5:eta1-cyclopentadienyl bridging ligands. syntheses and x-ray crystal structure of [(eta5-c5h5)m(mu-h)2(mu-(eta5:eta1-c5h4))irh(pmeph2)2][bph4] (m = mo and w). *Organometallics*, 5(9):1785–1791, September 1986. WOS:A1986E021400005.

[8] P. Amara, A. Volbeda, J. C. Fontecilla-Camps, and M. J. Field. A hybrid density functional theory molecular mechanics study of nickel-iron hydrogenase: Investigation of the active site redox states. *Journal of the American Chemical Society*, 121(18):4468–4477, May 1999. WOS:000080550100020.

[9] D. ANDRAE, U. HAUSSERMANN, M. DOLG, H. STOLL, and H. PREUSS. Energyadjusted abinitio pseudopotentials for the first row transition elements. *Theoretica Chimica Acta*, 77(2):123–141, April 1990. WOS:A1990DB14000004.

[10] Claudia Andreini, Ivano Bertini, Gabriele Cavallaro, Gemma L. Holliday, and Janet M. Thornton. Metal ions in biological catalysis: from enzyme databases to general principles. *JBIC Journal of Biological Inorganic Chemistry*, 13(8):1205–1218, November 2008.

[11] Claudia Andreini, Ivano Bertini, Gabriele Cavallaro, Gemma L. Holliday, and Janet M. Thornton. Metal-MACiE: a database of metals involved in biological catalysis. *Bioinformatics*, 25(16):2088–2089, August 2009.

[12] Claudia Andreini, Ivano Bertini, and Antonio Rosato. Metalloproteomes: A bioinformatic approach. *Accounts of Chemical Research*, 42(10):1471–1479, October 2009.

[13] Claudia Andreini, Gabriele Cavallaro, Serena Lorenzini, and Antonio Rosato. MetalPDB: a database of metal sites in biological macromolecular structures. *Nucleic Acids Research*, 41(D1):D312–D319, January 2013.

[14] Krassimira Angelova, Angelo Felline, Moon Lee, Manish Patel, David Puett, and Francesca Fanelli. Conserved amino acids participate in the structure networks deputed to intramolecular communication in the lutropin receptor. *Cellular and Molecular Life Sciences*, 68(7):1227–1239, April 2011.

[15] M. Aresta, A. Dibenedetto, and A Angelini. *CO_2 Chemistry: Converting "exhaust" carbon to "working" carbon.* Academic Press, 2013.

[16] X. Arias-Moreno, O. Abian, S. Vega, J. Sancho, and A. Velazquez-Campoy. Protein-cation interactions: structural and thermodynamic aspects. *Current Protein & Peptide Science*, 12(4):325–338, June 2011.

[17] Canan Atilgan, Osman Burak Okan, and Ali Rana Atilgan. Network-based models as tools hinting at nonevident protein functionality. *Annual Review of Biophysics*, 41(1):205–225, 2012.

[18] David S. Auld. Zinc coordination sphere in biochemical zinc sites. *Biometals*, 14(3-4):271–313, September 2001.

[19] Ayse Ozlem Aykut, Ali Rana Atilgan, and Canan Atilgan. Designing molecular dynamics simulations to shift populations of the conformational states of calmodulin. *PLoS Computational Biology*, 9(12), December 2013.

[20] Ivet Bahar, Timothy R. Lezon, Lee-Wei Yang, and Eran Eyal. Global dynamics of proteins: Bridging between structure and function. *Annual Review of Biophysics*, 39(1):23–42, 2010.

[21] R. Thomas Baker and William Tumas. Toward greener chemistry. *Science*, 284(5419):1477–1479, May 1999.

[22] Riccardo Baron, Sergio E. Wong, Cesar A. F. de Oliveira, and J. Andrew McCammon. E9-im9 colicin DNase-immunity protein biomolecular association in water: A multiple-copy and accelerated molecular dynamics simulation study. *The Journal of Physical Chemistry. B*, 112(51):16802–16814, December 2008.

[23] AD BECKE. density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, 38(6):3098–3100, September 1988. WOS:A1988Q146900044.

[24] AD BECKE. density-functional thermochemistry .1. the effect of the exchange-only gradient. *Journal of Chemical Physics*, 96(3):2155–2160, February 1992. WOS:A1992HC61000055.

[25] AD BECKE. density-functional thermochemistry .3. the role of exact exchange. *Journal of Chemical Physics*, 98(7):5648–5652, April 1993. WOS:A1993KV99700048.

[26] Moshe Ben-David, Grzegorz Wieczorek, Mikael Elias, Israel Silman, Joel L. Sussman, and Dan S. Tawfik. Catalytic metal ion rearrangements underline promiscuity and evolvability of a metalloenzyme. *Journal of Molecular Biology*, 425(6):1028–1038, March 2013.

[27] Eric E. Benson, Clifford P. Kubiak, Aaron J. Sathrum, and Jonathan M. Smieja. Electrocatalytic and homogeneous approaches to conversion of CO2 to liquid fuels. *Chemical Society Reviews*, 38(1):89–99, 2009. WOS:000261768900009.

[28] Gira Bhabha, Jeeyeon Lee, Damian C. Ekiert, Jongsik Gam, Ian A. Wilson, H. Jane Dyson, Stephen J. Benkovic, and Peter E. Wright. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science*, 332(6026):234–238, April 2011.

[29] Andrew N. Bigley and Frank M. Raushel. Catalytic mechanisms for phosphotriesterases. *Biochimica et biophysica acta*, 1834(1):443–453, January 2013.

[30] David D. Boehr, Ruth Nussinov, and Peter E. Wright. The role of dynamic conformational ensembles in biomolecular recognition. *Nature chemical biology*, 5(11):789–796, November 2009.

[31] David D. Boehr, Jason R. Schnell, Dan McElheny, Sung-Hun Bae, Brendan M. Duggan, Stephen J. Benkovic, H. Jane Dyson, and Peter E. Wright. A distal mutation perturbs dynamic amino acid networks in dihydrofolate reductase. *Biochemistry*, 52(27):4605–4619, July 2013.

[32] J. C. Boyington, V. N. Gladyshev, S. V. Khangulov, T. C. Stadtman, and P. D. Sun. Crystal structure of formate dehydrogenase h: Catalysis involving mo, molybdopterin, selenocysteine, and an fe4s4 cluster. *Science*, 275(5304):1305–1308, February 1997. WOS:A1997WK64400042.

[33] R. Breslow. Biomimetic chemistry: Biology as an inspiration. *Journal of Biological Chemistry*, 284(3):1337–1342, 2009.

[34] K. V. Brinda and Saraswathi Vishveshwara. A network representation of protein structures: Implications for protein stability. *Biophysical Journal*, 89(6):4159–4170, December 2005.

[35] Carlos D Brondino, Maria João Romão, Isabel Moura, and José JG Moura. Molybdenum and tungsten enzymes: the xanthine oxidase family. *Current Opinion in Chemical Biology*, 10(2):109–114, April 2006.

[36] M. Bruschi, L. De Gioia, G. Zampella, M. Reiher, P. Fantucci, and M. Stein. A theoretical study of spin states in ni-s-4 complexes and models of the [NiFe] hydrogenase active site. *Journal of Biological Inorganic Chemistry*, 9(7):873–884, October 2004. WOS:000224751400009.

[37] M. Bruschi, G. Zampella, P. Fantucci, and L. De Gioia. DFT investigations of models related to the active site of [NiFe] and [fe] hydrogenases. *Coordination Chemistry Reviews*, 249(15-16):1620–1640, August 2005. WOS:000231587200008.

[38] Pietro Buffa, Chiara Romano, Alessandro Pandini, Michele Massimino, Elena Tirrò, Francesco Di Raimondo, Livia Manzella, Franca Fraternali, and Paolo G. Vigneri. BCR-ABL residues interacting with ponatinib are critical to preserve the tumorigenic potential of the oncoprotein. *The FASEB Journal*, 28(3):1221–1236, March 2014.

[39] TE BURROW, A. HILLS, DL HUGHES, JD LANE, NJ LAZAROWYCH, MJ MAGUIRE, RH MORRIS, and RL RICHARDS. Cleavage of an aryl carbon–sulfur bond in hydride–thiolate complexes of molybdenum and tungsten. crystal structures of [Mo(SC6H2Pri3-2,4,6)(OMe)(PMePh2)2(µ-s)2] and [Mo(SC6H2Pri3-2,4,6)(OEt)(PEtPh2)2(µ-s)2]. *Journal of the Chemical Society-Chemical Communications*, (24):1757–1759, December 1990. WOS:A1990ET45100008.

[40] Csaba Böde, István A. Kovács, Máté S. Szalay, Robin Palotai, Tamás Korcsmáros, and Péter Csermely. Network analysis of protein dynamics. *FEBS Letters*, 581(15):2776–2782, June 2007.

[41] Sigolene Canaguier, Martin Field, Yohan Oudart, Jacques Pecaut, Marc Fontecave, and Vincent Artero. A structural and functional mimic of the active site of NiFe hydrogenases. *Chemical Communications*, 46(32):5876–5878, 2010. WOS:000280658300008.

[42] Maria E. Carroll, Bryan E. Barton, Danielle L. Gray, Amanda E. Mack, and Thomas B. Rauchfuss. Active-site models for the nickel-iron hydrogenases: Effects of ligands on reactivity and catalytic properties. *Inorganic Chemistry*, 50(19):9554–9563, October 2011. WOS:000295115000044.

[43] Chin Jung Cheng and Valerie Daggett. Different misfolding mechanisms converge on common conformational changes: Human prion

protein pathogenic mutants y218n and e196k. *Prion*, 8(1):125–135, February 2014.

[44] Fernanda Cimadevilla, M. Esther Garcia, Daniel Garcia-Vivo, Miguel A. Ruiz, M. Teresa Rueda, and Sabine Halut. Protonation reactions of the oxo complex cis-[mo-2(eta(5)-c5h5)(2)(o)(mu-PPh2)(2)(CO)]. hydroxo and tetrafluoroborate derivatives. *Journal of Organometallic Chemistry*, 699:67–74, February 2012. WOS:000299118700010.

[45] Joseph E Coleman. Zinc enzymes. *Current Opinion in Chemical Biology*, 2(2):222–234, April 1998.

[46] Dario Corrada and Giorgio Colombo. Energetic and dynamic aspects of the affinity maturation process: Characterizing improved variants from the bevacizumab antibody with molecular simulations. *Journal of Chemical Information and Modeling*, 53(11):2937–2950, November 2013.

[47] Robert R Crichton. *Biological inorganic chemistry*. Elsevier, 2011 edition, 2011.

[48] Peter Csermely, Tamas Korcsmaros, Huba J. M. Kiss, Gabor London, and Ruth Nussinov. Structure and dynamics of molecular networks: A novel paradigm of drug discovery. a comprehensive review. *Pharmacology & Therapeutics*, 138(3):333–408, June 2013. arXiv: 1210.0330.

[49] J. R. R. Fausto Da Silva and R. J. P. Williams. *The biological chemistry of the elements*. Oxford Press, second edition, 2001.

[50] Ilda D'Annessa, Cinzia Tesauro, Zhenxing Wang, Barbara Arnò, Laura Zuccaro, Paola Fiorani, and Alessandro Desideri. The human topoisomerase 1b arg634ala mutation results in camptothecin resistance and loss of inter-domain motion correlation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1834(12):2712–2721, December 2013.

[51] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An nlog(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, June 1993.

[52] DJ DARENSBOURG, MB FISCHER, RE SCHMIDT, and BJ BALDWIN. Formate ion as a monodentate ligand. synthesis, structure, and decarboxylation of (.eta.5-cyclopentadienyl)dicarbonyl(formato)iron. *Journal of the American Chemical Society*, 103(5):1297–1298, 1981. WOS:A1981LF69400087.

[53] DJ DARENSBOURG, A. ROKICKI, and MY DARENSBOURG. Facile reduction of carbon dioxide by anionic group 6b metal hydrides. chemistry relevant to catalysis of the water-gas shift reaction. *Journal of the American Chemical Society*, 103(11):3223–3224, 1981. WOS:A1981LS85100055.

[54] A. L. De Lacey, A. Pardo, V. M. Fernandez, S. Dementin, G. Adryanczyk-Perrier, E. C. Hatchikian, and M. Rousset. FTIR spectroelectrochemical study of the activation and inactivation processes of [NiFe] hydrogenases: effects of solvent isotope replacement and site-directed mutagenesis. *Journal of Biological Inorganic Chemistry*, 9(5):636–642, July 2004. WOS:000223014000015.

[55] Antonio L. De Lacey, Víctor M. Fernández, Marc Rousset, and Richard Cammack. Activation and inactivation of hydrogenase function and the catalytic cycle: spectroelectrochemical studies. *Chemical Reviews*, 107(10):4304–4330, 2007. PMID: 17715982.

[56] Antonio del Sol, Chung-Jung Tsai, Buyong Ma, and Ruth Nussinov. The origin of allosteric functional modulation: Multiple pre-existing pathways. *Structure (London, England : 1993)*, 17(8):1042–1050, August 2009.

[57] S. Dementin, B. Burlat, A. L. De Lacey, A. Pardo, G. Adryanczyk-Perrier, B. Guigliarelli, V. M. Fernandez, and M. Rousset. A glutamate is the essential proton transfer gate during the catalytic cycle of the [NiFe] hydrogenase. *Journal of Biological Chemistry*, 279(11):10508–10513, March 2004. WOS:000220050400103.

[58] Guy G Dodson, David P Lane, and Chandra S Verma. Molecular simulations of protein dynamics: new windows on mechanisms in biology. *EMBO Reports*, 9(2):144–150, February 2008.

[59] F. Dole, A. Fournel, V. Magro, E. C. Hatchikian, P. Bertrand, and B. Guigliarelli. Nature and electronic structure of the ni-x dinuclear center of desulfovibrio gigas hydrogenase. implications for the

enzymatic mechanism. *Biochemistry*, 36(25):7847–7854, June 1997. WOS:A1997XG90100026.

[60] Urmi Doshi, Lauren C. McGowan, Safieh Tork Ladani, and Donald Hamelberg. Resolving the complex role of enzyme conformational dynamics in catalytic function. *Proceedings of the National Academy of Sciences*, 109(15):5699–5704, April 2012.

[61] Ron O. Dror, Robert M. Dirks, J.P. Grossman, Huafeng Xu, and David E. Shaw. Biomolecular simulation: A computational microscope for molecular biology. *Annual Review of Biophysics*, 41(1):429–452, 2012.

[62] Daniel L. DuBois, Daniel M. Blake, Alex Miedaner, Calvin J. Curtis, M. R. DuBois, James A. Franz, and John C. Linehan. Hydride transfer from rhodium complexes to triethylborane. *Organometallics*, 25(18):4414–4419, August 2006. WOS:000239887600023.

[63] Todor Dudev. Modeling metal binding sites in proteins by quantum chemical calculations. *Modeling Metal Binding Sites in Proteins by Quantum Chemical Calculations*, 2:19–21, 2014.

[64] K. EICHKORN, O. TREUTLER, H. OHM, M. HASER, and R. AHLRICHS. auxiliary basis sets to approximate coulomb potentials. *Chemical Physics Letters*, 240(4):283–289, June 1995. WOS:A1995RG42300007.

[65] Richard Eisenberg and Dan E. Hendriksen. *The Binding and Activation of Carbon Monoxide, Carbon Dioxide, and Nitric Oxide and Their Homogeneously Catalyzed Reactions*, volume 28 of *Advances in Catalysis*. Academic Press, 1979.

[66] Elan Z. Eisenmesser, Oscar Millet, Wladimir Labeikovsky, Dmitry M. Korzhnev, Magnus Wolf-Watz, Daryl A. Bosco, Jack J. Skalicky, Lewis E. Kay, and Dorothee Kern. Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, 438(7064):117–121, November 2005.

[67] Elan Zohar Eisenmesser, Daryl A. Bosco, Mikael Akke, and Dorothee Kern. Enzyme dynamics during catalysis. *Science*, 295(5559):1520–1523, February 2002.

[68] F. Ely, J. L. Foo, C. J. Jackson, L. R. Gahan, D. L. Ollis, and G. Schenk. Enzymatic bioremediation: Organophosphate degradation by binuclear metallo-hydrolases. *Current Topics in Biochemical Research*, 2008(9):63–78, 2008.

[69] Niall J. English, Morad M. El-Hendawy, Damian A. Mooney, and J. M. D. MacElroy. Perspectives on atmospheric CO2 fixation in inorganic and biomimetic structures. *Coordination Chemistry Reviews*, 269:85–95, June 2014.

[70] Ernesto Estrada. Universality in protein residue networks. *Biophysical Journal*, 98(5):890–900, March 2010.

[71] Kramer T Krampa M Lubitz W Neese F. Theoretical spectroscopy of the ni(ii) intermediate states in the catalytic cycle and the activation of [nife] hydrogenases. *Chembiochem*, 14(14):1898–1905, September 2013.

[72] Marc Fontecave and Vincent Artero. Bioinspired catalysis at the crossroads between biology and chemistry: A remarkable example of an electrocatalytic material mimicking hydrogenases. *Comptes Rendus Chimie*, 14(4):362–371, April 2011.

[73] Juan C. Fontecilla-Camps, Patricia Amara, Christine Cavazza, Yvain Nicolet, and Anne Volbeda. Structure-function relationships of anaerobic gas-processing metalloenzymes. *Nature*, 460(7257):814–822, August 2009. WOS:000268938300025.

[74] Juan C. Fontecilla-Camps, Patricia Amara, Christine Cavazza, Yvain Nicolet, and Anne Volbeda. Structure–function relationships of anaerobic gas-processing metalloenzymes. *Nature*, 460(7257):814–822, August 2009.

[75] 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, October 2007.

[76] 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, October 2007. WOS:000249839900013.

[77] JeeLoon Foo, ColinJ. Jackson, PaulD. Carr, HyeKyung Kim, Gerhard Schenk, LawrenceR. Gahan, and DavidL. Ollis. Mutation of outer-shell residues modulates metal ion co-ordination strength in a metalloenzyme. *Biochemical Journal*, 429(2):313–321, July 2010.

[78] Vincent Fourmond, Sigolene Canaguier, Benjamin Golly, Martin J. Field, Marc Fontecave, and Vincent Artero. A nickel-manganese catalyst as a biomimic of the active site of NiFe hydrogenases: a combined electrocatalytical and DFT mechanistic study. *Energy & Environmental Science*, 4(7):2417–2427, July 2011. WOS:000292205100010.

[79] James S. Fraser, Michael W. Clarkson, Sheena C. Degnan, Renske Erion, Dorothee Kern, and Tom Alber. Hidden alternate structures of proline isomerase essential for catalysis. *Nature*, 462(7273):669–673, December 2009.

[80] Filip Fratev, Elina Mihaylova, and Ilza Pajeva. Combination of genetic screening and molecular dynamics as a useful tool for identification of disease-related mutations: ZASP PDZ domain g54s mutation case. *Journal of Chemical Information and Modeling*, 54(5):1524–1536, May 2014.

[81] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian09 Revision D.01. Gaussian Inc. Wallingford CT 2009.

[82] Johannes Fritsch, Patrick Scheerer, Stefan Frielingsdorf, Sebastian Kroschinsky, Baerbel Friedrich, Oliver Lenz, and Christian M. T. Spahn. The crystal structure of an oxygen-tolerant hydrogenase

uncovers a novel iron-sulphur centre. *Nature*, 479(7372):249–U134, November 2011. WOS:000298030800047.

[83] F. Galan, M. Fouassier, M. Tranquille, J. Mascetti, and I. Papai. CO_2 coordination to nickel atoms: Matrix isolation and density functional studies. *Journal of Physical Chemistry A*, 101(14):2626–2633, April 1997. WOS:A1997WR42700017.

[84] Ignacio Fdez Galvan, Anne Volbeda, Juan C. Fontecilla-Camps, and Martin J. Field. A QM/MM study of proton transport pathways in a [NiFe] hydrogenase. *Proteins-Structure Function and Bioinformatics*, 73(1):195–203, October 2008. WOS:000259114500019.

[85] Ibram Ganesh. Conversion of carbon dioxide into methanol – a potential liquid fuel: Fundamental challenges and opportunities (a review). *Renewable and Sustainable Energy Reviews*, 31:221–257, March 2014.

[86] Mireia Garcia-Viloca, Jiali Gao, Martin Karplus, and Donald G. Truhlar. How enzymes work: Analysis by modern rate theory and computer simulations. *Science*, 303(5655):186–195, January 2004.

[87] I. P. Georgakaki, L. M. Thomson, E. J. Lyon, M. B. Hall, and M. Y. Darensbourg. Fundamental properties of small molecule models of fe-only hydrogenase: computations relative to the definition of an entatic state in the active site. *Coordination Chemistry Reviews*, 238:255–266, March 2003. WOS:000182957600014.

[88] Amit Ghosh, Reiko Sakaguchi, Cuiping Liu, Saraswathi Vishveshwara, and Ya-Ming Hou. Allosteric communication in cysteinyl tRNA synthetase. *The Journal of Biological Chemistry*, 286(43):37721–37731, October 2011.

[89] Amit Ghosh and Saraswathi Vishveshwara. A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 104(40):15711–15716, October 2007.

[90] Lesley H. Greene and Victoria A. Higman. Uncovering network systems within protein structures. *Journal of Molecular Biology*, 334(4):781–791, December 2003.

[91] S. Grimme. Accurate description of van der waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry*, 25(12):1463–1473, September 2004. WOS:000222832100004.

[92] Stefan Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry*, 27(15):1787–1799, November 2006. WOS:000241477200003.

[93] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-d) for the 94 elements h-pu. *Journal of Chemical Physics*, 132(15):154104, April 2010. WOS:000276971500005.

[94] Gordon G. Hammes, Stephen J. Benkovic, and Sharon Hammes-Schiffer. Flexibility, diversity, and cooperativity: Pillars of enzyme catalysis. *Biochemistry*, 50(48):10422–10430, December 2011.

[95] R. P. Happe, W. Roseboom, A. J. Pierik, S. P. J. Albracht, and K. A. Bagley. Biological activation of hydrogen. *Nature*, 385(6612):126–126, January 1997. WOS:A1997WB72800037.

[96] Marjorie M. Harding, Matthew W. Nowicki, and Malcolm D. Walkinshaw. Metals in protein structures: a review of their principal features. *Crystallography Reviews*, 16(4):247–302, 2010.

[97] T. HELGAKER. Transition-state optimizations by trust-region image minimization. *Chemical Physics Letters*, 182(5):503–510, August 1991. WOS:A1991GB80500021.

[98] Katherine A. Henzler-Wildman, Ming Lei, Vu Thai, S. Jordan Kerns, Martin Karplus, and Dorothee Kern. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, 450(7171):913–916, December 2007.

[99] Katherine A. Henzler-Wildman, Ming Lei, Vu Thai, S. Jordan Kerns, Martin Karplus, and Dorothee Kern. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, 450(7171):913–916, December 2007.

[100] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, September 1997.

[101] Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, March 2008.

[102] Suzannah V. Hexter, Felix Grey, Thomas Happe, Victor Climent, and Fraser A. Armstrong. Electrocatalytic mechanism of reversible hydrogen cycling by enzymes and distinctions between the major classes of hydrogenases. *Proceedings of the National Academy of Sciences of the United States of America*, 109(29):11516–11521, July 2012. WOS:000306837100015.

[103] Y. Higuchi, H. Ogata, K. Miki, N. Yasuoka, and T. Yagi. Removal of the bridging ligand atom at the ni-fe active site of [NiFe] hydrogenase upon reduction with h-2, as revealed by x-ray structure analysis at 1.4 angstrom resolution. *Structure with Folding & Design*, 7(5):549–556, May 1999. WOS:000080577000012.

[104] Russ Hille. The mononuclear molybdenum enzymes†. *Chemical Reviews*, 96(7):2757–2816, January 1996.

[105] Russ Hille, James Hall, and Partha Basu. The mononuclear molybdenum enzymes. *Chemical Reviews*, 114(7):3963–4038, April 2014.

[106] Vincent J. Hilser, Bertrand García-Moreno E., Terrence G. Oas, Greg Kapp, and Steven T. Whitten. A statistical thermodynamic model of the protein ensemble. *Chemical Reviews*, 106(5):1545–1558, May 2006.

[107] J. D. Holladay, J. Hu, D. L. King, and Y. Wang. An overview of hydrogen production technologies. *Catalysis Today*, 139(4):244–260, January 2009.

[108] Richard H. Holm, Pierre Kennepohl, and Edward I. Solomon. Structural and functional aspects of metal sites in biology. *Chemical Reviews*, 96(7):2239–2314, November 1996.

[109] F. Galan J. Mascetti and I. Pàpai. Carbon dioxide interaction with metal atoms: matrix isolation spectroscopic study and {DFT} calculations. *Coordination Chemistry Reviews*, 190–192(0):557 – 576, 1999.

[110] C. J. Jackson, J.-L. Foo, N. Tokuriki, L. Afriat, P. D. Carr, H.-K. Kim, G. Schenk, D. S. Tawfik, and D. L. Ollis. Conformational sampling, catalysis, and evolution of the bacterial phosphotriesterase. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51):21631–21636, December 2009.

[111] Gonzalo Jiménez-Osés, Sílvia Osuna, Xue Gao, Michael R. Sawaya, Lynne Gilson, Steven J. Collier, Gjalt W. Huisman, Todd O. Yeates, Yi Tang, and K. N. Houk. The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nature Chemical Biology*, 10(6):431–436, June 2014.

[112] Deborah C. Johnson, Dennis R. Dean, Archer D. Smith, and Michael K. Johnson. Structure, function, and formation of biological iron-sulfur clusters. *Annual Review of Biochemistry*, 74(1):247–281, 2005.

[113] Kabsch. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12), 1983.

[114] Shina C. L. Kamerlin and Arieh Warshel. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins: Structure, Function, and Bioinformatics*, pages NA–NA, 2010.

[115] Mario Kampa, Wolfgang Lubitz, Maurice van Gastel, and Frank Neese. Computational study of the electronic structure and magnetic properties of the ni-c state in [NiFe] hydrogenases including the second coordination sphere. *Journal of Biological Inorganic Chemistry*, 17(8):1269–1281, December 2012. WOS:000311669200012.

[116] M. Karplus and J. Kuriyan. Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences of the United States of America*, 102(19):6679–6685, May 2005.

[117] Jason M. Keith and Michael B. Hall. Potential hydrogen bottleneck in nickel-iron hydrogenase. *Inorganic Chemistry*, 49(14):6378–6380, July 2010. WOS:000279621200016.

[118] S. V. Khangulov, V. N. Gladyshev, G. C. Dismukes, and T. C. Stadtman. Selenium-containing formate dehydrogenase h from escherichia coli: A molybdopterin enzyme that catalyzes formate oxidation without oxygen transfer. *Biochemistry*, 37(10):3518–3528, March 1998. WOS:000072630300033.

[119] A. KLAMT. Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena. *Journal of Physical Chemistry*, 99(7):2224–2235, February 1995. WOS:A1995QG92400062.

[120] A. Klamt and G. Schüürmann. Cosmo: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2*, pages 799–805, 1993.

[121] 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):120–127, April 2009.

[122] Elena G. Kovaleva and John D. Lipscomb. Versatility of biological non-heme fe(II) centers in oxygen activation reactions. *Nature Chemical Biology*, 4(3):186–193, March 2008.

[123] AT KOWAL, IC ZAMBRANO, I. MOURA, JJG MOURA, J. LEGALL, and MK JOHNSON. Electronic and magnetic properties of nickel-substituted rubredoxin: a variable-temperature magnetic circular dichroism study. *Inorganic Chemistry*, 27(7):1162–1166, April 1988. WOS:A1988M920300016.

[124] Sebastian Kozuch and Sason Shaik. How to conceptualize catalytic cycles? the energetic span model. *Accounts of Chemical Research*, 44(2):101–110, February 2011.

[125] Ambuj Kumar and Rituraj Purohit. Use of long term molecular dynamics simulation in predicting cancer associated SNPs. *PLoS Computational Biology*, 10(4), April 2014.

[126] Elodie Laine, Christian Auclair, and Luba Tchertanov. Allosteric communication across the native and mutated KIT receptor tyrosine kinase. *PLoS Computational Biology*, 8(8), August 2012.

[127] Oliver F. Lange and Helmut Grubmüller. Generalized correlation for biomolecular dynamics. *Proteins: Structure, Function, and Bioinformatics*, 62(4):1053–1061, March 2006.

[128] CT LEE, WT YANG, and RG PARR. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2):785–789, January 1988. WOS:A1988L976200011.

[129] Monica Leopoldini, Sandro G. Chiodo, Marirosa Toscano, and Nino Russo. Reaction mechanism of molybdoenzyme formate dehydrogenase. *Chemistry-a European Journal*, 14(28):8674–8681, 2008. WOS:000260035900029.

[130] Kimfung Li, Xiaoqiang An, Kyeong Hyeon Park, Majeda Khraisheh, and Junwang Tang. A critical review of CO_2 photoconversion: Catalysts and reactors. *Catalysis Today*, 224:3–12, April 2014.

[131] Sten O. Nilsson Lill and Per E. M. Siegbahn. An autocatalytic mechanism for NiFe-hydrogenase: Reduction to ni(i) followed by oxidative addition. *Biochemistry*, 48(5):1056–1066, February 2009. WOS:000263047900027.

[132] Rern Jern Lim, Mingshi Xie, Mahasin Alam Sk, Jong-Min Lee, Adrian Fisher, Xin Wang, and Kok Hwa Lim. A review on the electrochemical reduction of CO_2 in fuel cells, metal electrodes and molecular catalysts. *Catalysis Today*, 233:169–180, September 2014.

[133] Stephen J. Lippard. The inorganic side of chemical biology. *Nature Chemical Biology*, 2(10):504–507, October 2006.

[134] Xu Lu, Dennis Y. C. Leung, Huizhi Wang, Michael K. H. Leung, and Jin Xuan. Electrochemical reduction of carbon dioxide to formic acid. *ChemElectroChem*, 1(5):836–849, May 2014.

[135] Marcus Ludwig, James A. Cracknell, Kylie A. Vincent, Fraser A. Armstrong, and Oliver Lenz. Oxygen-tolerant h-2 oxidation by membrane-bound [NiFe] hydrogenases of ralstonia species coping with low level h2 in air. *Journal of Biological Chemistry*, 284(1):465–477, January 2009. WOS:000261974800051.

[136] Suryani Lukman, Barry J. Grant, Alemayehu A. Gorfe, Guy H. Grant, and J. Andrew McCammon. The distinct conformational

dynamics of k-ras and h-ras a59g. *PLoS Computational Biology*, 6(9), September 2010.

[137] Simona Mariani, Daniele Dell'Orco, Angelo Felline, Francesco Raimondi, and Francesca Fanelli. Network and atomistic simulations unveil the structural determinants of mutations linked to retinal diseases. *PLoS Computational Biology*, 9(8), August 2013.

[138] Kelly G. Matz, Regina P. Mtei, Rebecca Rothstein, Martin L. Kirk, and Sharon J. Nieter Burgmayer. Study of molybdenum(4+) quinoxalyldithiolenes as models for the noninnocent pyranopterin in the molybdenum cofactor. *Inorganic Chemistry*, 50(20):9804–9815, October 2011. WOS:000295602500007.

[139] Alastair G. McEwan, Justin P. Ridge, Christopher A. McDevitt, and Philip Hugenholtz. The DMSO reductase family of microbial molybdenum enzymes; molecular properties and role in the dissimilatory reduction of toxic elements. *Geomicrobiology Journal*, 19(1):3–21, January 2002.

[140] Carlo Mealli and Thomas B. Rauchfuss. Models for the hydrogenases put the focus where it should be-hydrogen. *Angewandte Chemie-International Edition*, 46(47):8942–8944, 2007. WOS:000251576400002.

[141] Lidio Meireles, Mert Gur, Ahmet Bakan, and Ivet Bahar. Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins. *Protein Science : A Publication of the Protein Society*, 20(10):1645–1658, October 2011.

[142] Mark F. Mohamed and Florian Hollfelder. Efficient, crosswise catalytic promiscuity among enzymes that catalyze phosphoryl transfer. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1834(1):417–424, January 2013.

[143] Y. Montet, P. Amara, A. Volbeda, X. Vernede, E. C. Hatchikian, M. J. Field, M. Frey, and J. C. FontecillaCamps. Gas access to the active site of ni-fe hydrogenases probed by x-ray crystallography and molecular dynamics. *Nature Structural Biology*, 4(7):523–526, July 1997. WOS:A1997XH73500008.

[144] Cristiano S. Mota, Maria G. Rivas, Carlos D. Brondino, Isabel Moura, Jose J. G. Moura, Pablo J. Gonzalez, and Nuno M. F. S. A. Cerqueira. The mechanism of formate oxidation

by metal-dependent formate dehydrogenases. *Journal of Biological Inorganic Chemistry*, 16(8):1255–1268, December 2011. WOS:000297227000012.

[145] Vishal C Nashine, Sharon Hammes-Schiffer, and Stephen J Benkovic. Coupled motions in enzyme catalysis. *Current Opinion in Chemical Biology*, 14(5):644–651, October 2010.

[146] R. M. Navarro, M. A. Peña, and J. L. G. Fierro. Hydrogen production reactions from carbon feedstocks: fossil fuels and biomass. *Chemical Reviews*, 107(10):3952–3991, October 2007.

[147] Y. Nicolet, C. Cavazza, and J. C. Fontecilla-Camps. Fe-only hydrogenases: structure, function and evolution. *Journal of Inorganic Biochemistry*, 91(1):1–8, July 2002. WOS:000177141500002.

[148] S. Niu and N. B. Hall. Modeling the active sites in metalloenzymes 5. the heterolytic bond cleavage of h-2 in the [NiFe] hydrogenase of desulfovibrio gigas by a nucleophilic addition mechanism. *Inorganic Chemistry*, 40(24):6201–6203, November 2001. WOS:000172181400018.

[149] S. Q. Niu, L. M. Thomson, and M. B. Hall. Theoretical characterization, of the reaction intermediates in a model of the nickel-iron hydrogenase of desulfovibrio gigas. *Journal of the American Chemical Society*, 121(16):4000–4007, April 1999. WOS:000080181100022.

[150] Hideaki Ogata, Petra Kellers, and Wolfgang Lubitz. The crystal structure of the [NiFe] hydrogenase from the photosynthetic bacterium allochromatium vinosum: Characterization of the oxidized enzyme (ni-a state). *Journal of Molecular Biology*, 402(2):428–444, September 2010. WOS:000283208700011.

[151] Hideaki Ogata, Wolfgang Lubitz, and Yoshiki Higuchi. [nife] hydrogenases: structural and spectroscopic studies of the reaction mechanism. *Dalton Trans.*, pages 7577–7587, 2009.

[152] Seiji Ogo, Koji Ichikawa, Takahiro Kishima, Takahiro Matsumoto, Hidetaka Nakai, Katsuhiro Kusaka, and Takashi Ohhara. A functional [NiFe]hydrogenase mimic that catalyzes electron and hydride transfer from h-2. *Science*, 339(6120):682–684, February 2013. WOS:000314585600038.

[153] Seiji Ogo, Ryota Kabe, Keiji Uehara, Bunsho Kure, Takashi Nishimura, Saija C. Menon, Ryosuke Harada, Shunichi Fukuzumi, Yoshiki Higuchi, Takashi Ohhara, Taro Tamada, and Ryota Kuroki. A dinuclear ni(mu-h)ru complex derived from h-2. *Science*, 316(5824):585–587, April 2007. WOS:000245983100041.

[154] Maria-Eirini Pandelia, Hideaki Ogata, and Wolfgang Lubitz. Intermediates in the catalytic cycle of [NiFe] hydrogenase: Functional spectroscopy of the active site. *Chemphyschem*, 11(6):1127–1140, April 2010. WOS:000277666900004.

[155] Alessandro Pandini, Arianna Fornili, Franca Fraternali, and Jens Kleinjung. Detection of allosteric signal transmission by information-theoretic analysis of protein dynamics. *The FASEB Journal*, 26(2):868–881, February 2012.

[156] Elena Papaleo, Kresten Lindorff-Larsen, and Luca De Gioia. Paths of long-range communication in the e2 enzymes of family 3: a molecular dynamics investigation. *Physical Chemistry Chemical Physics*, 14(36):12515–12525, August 2012.

[157] Elena Papaleo, Kresten Lindorff-Larsen, and Luca De Gioia. Paths of long-range communication in the e2 enzymes of family 3: a molecular dynamics investigation. *Physical Chemistry Chemical Physics*, 14(36):12515–12525, August 2012.

[158] Elena Papaleo, Giulia Renzetti, and Matteo Tiberti. Mechanisms of intramolecular communication in a hyperthermophilic acylaminoacyl peptidase: A molecular dynamics investigation. *PLoS ONE*, 7(4):e35686, April 2012.

[159] A. Pardo, A. L. De Lacey, V. M. Fernandez, H. J. Fan, Y. B. Fan, and M. B. Hall. Density functional study of the catalytic cycle of nickel-iron [NiFe] hydrogenases and the involvement of high-spin nickel(II). *Journal of Biological Inorganic Chemistry*, 11(3):286–306, April 2006. WOS:000236586000004.

[160] Alison Parkin and Frank Sargent. The hows and whys of aerobic h-2 metabolism. *Current Opinion in Chemical Biology*, 16(1-2):26–34, April 2012. WOS:000303640400005.

[161] Marco Pasi, Matteo Tiberti, Alberto Arrigoni, and Elena Papaleo. xPyder: A PyMOL plugin to analyze coupled residues and their networks in protein structures. *Journal of Chemical Information and Modeling*, 52(7):1865–1874, July 2012.

[162] JP PERDEW. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Physical Review B*, 33(12):8822–8824, June 1986. WOS:A1986C899500063.

[163] Stefano Piana, Kresten Lindorff-Larsen, and DavidE. Shaw. How robust are protein folding simulations with respect to force field parameterization? *Biophysical Journal*, 100(9):L47–L49, May 2011.

[164] A. J. Pierik, W. Roseboom, R. P. Happe, K. A. Bagley, and S. P. J. Albracht. Carbon monoxide and cyanide as intrinsic ligands to iron in the active site of [NiFe]-hydrogenases - NiFe(CN)(2)CO, biology's way to activate h-2. *Journal of Biological Chemistry*, 274(6):3331–3337, February 1999. WOS:000078428200014.

[165] Andrei V. Pisliakov, Jie Cao, Shina C. L. Kamerlin, and Arieh Warshel. Enzyme millisecond conformational dynamics do not catalyze the chemical step. *Proceedings of the National Academy of Sciences*, 106(41):17359–17364, October 2009.

[166] United Nations Environment Programme. Green economy report (http://www.unep.org/greeneconomy/GreenEconomyReport), 2011.

[167] Hans C. A. Raaijmakers and Maria Joao Romao. Formate-reduced e-coli formate dehydrogenase h: the reinterpretation of the crystal structure suggests a new reaction mechanism. *Journal of Biological Inorganic Chemistry*, 11(7):849–854, October 2006. WOS:000240283800005.

[168] J. W. Raebiger, A. Miedaner, C. J. Curtis, S. M. Miller, O. P. Anderson, and D. L. DuBois. Using ligand bite angles to control the hydricity of palladium diphosphine complexes. *Journal of the American Chemical Society*, 126(17):5502–5514, May 2004. WOS:000221135400037.

[169] Arvind Ramanathan and Pratul K. Agarwal. Evolutionarily conserved linkage between enzyme fold, flexibility, and catalysis. *PLoS Biol*, 9(11):e1001193, November 2011.

[170] Arvind Ramanathan, Andrej Savol, Virginia Burger, Chakra S. Chennubhotla, and Pratul K. Agarwal. Protein conformational populations and functionally relevant substates. *Accounts of Chemical Research*, 47(1):149–156, January 2014.

[171] Arvind Ramanathan, Andrej Savol, Virginia Burger, Chakra S. Chennubhotla, and Pratul K. Agarwal. Protein conformational populations and functionally relevant substates. *Accounts of Chemical Research*, 47(1):149–156, January 2014.

[172] AE REED and F. WEINHOLD. Natural bond orbital analysis of nearhartree–fock water dimer. *Journal of Chemical Physics*, 78(6):4066–4073, 1983. WOS:A1983QG96300090.

[173] M. Reiher, O. Salomon, and B. A. Hess. Reparameterization of hybrid functionals based on energy differences of states of different multiplicity. *Theoretical Chemistry Accounts*, 107(1):48–55, December 2001. WOS:000173203700008.

[174] Andre A. S. T. Ribeiro and Vanessa Ortiz. Determination of signaling pathways in proteins through network theory: Importance of the topology. *Journal of Chemical Theory and Computation*, 10(4):1762–1769, April 2014.

[175] Frederic Rousseau and Joost Schymkowitz. A systems biology perspective on protein structural dynamics and signal transduction. *Current Opinion in Structural Biology*, 15(1):23–30, February 2005.

[176] Amitava Roy and Carol Beth Post. Detection of long-range concerted motions in protein by a distance covariance. *Journal of Chemical Theory and Computation*, 8(9):3009–3014, September 2012.

[177] Ulf Ryde, Carola Schulzke, and Kerstin Starke. Which functional groups of the molybdopterin ligand should be considered when modeling the active sites of the molybdenum and tungsten cofactors? a density functional theory study. *Journal of Biological Inorganic Chemistry*, 14(7):1053–1064, September 2009. WOS:000269912600006.

[178] Sophie Sacquin-Mora, Pierre Sebban, Valérie Derrien, Bernhard Frick, Richard Lavery, and Christiane Alba-Simionesco. Probing the flexibility of the bacterial reaction center: the wild-type protein is more rigid than two site-specific mutants. *Biochemistry*, 46(51):14960–14968, December 2007.

[179] Miguel Saggu, Marcus Ludwig, Baerbel Friedrich, Peter Hildebrandt, Robert Bittl, Friedhelm Lendzian, Oliver Lenz, and Ingo Zebger. Impact of amino acid substitutions near the catalytic

site on the spectral properties of an o-2-tolerant membrane-bound [NiFe] hydrogenase. *Chemphyschem*, 11(6):1215–1224, April 2010. WOS:000277666900013.

[180] Maria Chiara Scaini, Giovanni Minervini, Lisa Elefanti, Paola Ghiorzo, Lorenza Pastorino, Silvia Tognazzo, Simona Agata, Monica Quaggio, Daniela Zullato, Giovanna Bianchi-Scarrà, Marco Montagna, Emma D'Andrea, Chiara Menin, and Silvio C.E. Tosatto. CDKN2a unclassified variants in familial malignant melanoma: Combining functional and computational approaches for their assessment. *Human Mutation*, 35(7):828–840, 2014.

[181] A. SCHAFER, C. HUBER, and R. AHLRICHS. Fully optimized contracted gaussian basis sets of triple zeta valence quality for atoms li to kr. *Journal of Chemical Physics*, 100(8):5829–5835, April 1994. WOS:A1994NF08100043.

[182] A. Schafer, A. Klamt, D. Sattel, J. C. W. Lohrenz, and F. Eckert. COSMO implementation in TURBOMOLE: Extension of an efficient quantum chemical code towards liquid systems. *Physical Chemistry Chemical Physics*, 2(10):2187–2193, 2000. WOS:000086928000018.

[183] David Schilter, Mark J. Nilges, Mrinmoy Chakrabarti, Paul A. Lindahl, Thomas B. Rauchfuss, and Matthias Stein. Mixed-valence nickel-iron dithiolate models of the [NiFe]-hydrogenase active site. *Inorganic Chemistry*, 51(4):2338–2348, February 2012. WOS:000300466300046.

[184] Steven D. Schwartz and Vern L. Schramm. Enzymatic transition states and dynamic motion in barrier crossing. *Nature Chemical Biology*, 5(8):551–558, August 2009.

[185] Hannah S. Shafaat, Olaf Rüdiger, Hideaki Ogata, and Wolfgang Lubitz. [NiFe] hydrogenases: A common active site for hydrogen metabolism under diverse conditions. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1827(8–9):986–1002, August 2013.

[186] Per E. M. Siegbahn and Fahmi Himo. Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *Journal of Biological Inorganic Chemistry*, 14(5):643–651, June 2009. WOS:000266924400001.

[187] Per E. M. Siegbahn and Fahmi Himo. The quantum chemical cluster approach for modeling enzyme reactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(3):323–336, June 2011.

[188] Per E. M. Siegbahn, Jesse W. Tye, and Michael B. Hall. Computational studies of [NiFe] and [FeFe] hydrogenases. *Chemical Reviews*, 107(10):4414–4435, October 2007. WOS:000249839900017.

[189] Yolanda A. Small, Daniel L. DuBois, Etsuko Fujita, and James T. Muckerman. Proton management as a design principle for hydrogenase-inspired catalysts. *Energy & Environmental Science*, 4(8):3008–3020, August 2011. WOS:000293213600050.

[190] Edward I. Solomon, Andrea Decker, and Nicolai Lehnert. Non-heme iron enzymes: Contrasts to heme catalysis. *Proceedings of the National Academy of Sciences*, 100(7):3589–3594, April 2003.

[191] Edward I. Solomon, Kenneth M. Light, Lei V. Liu, Martin Srnec, and Shaun D. Wong. Geometric and electronic structure contributions to function in non-heme iron enzymes. *Accounts of Chemical Research*, 46(11):2725–2739, November 2013.

[192] Susan Solomon, John S. Daniel, Todd J. Sanford, Daniel M. Murphy, Gian-Kasper Plattner, Reto Knutti, and Pierre Friedlingstein. Persistence of climate changes due to a range of greenhouse gases. *Proceedings of the National Academy of Sciences*, 107(43):18354–18359, October 2010.

[193] M. Stein and W. Lubitz. Relativistic DFT calculation of the reaction cycle intermediates of [NiFe] hydrogenase: a contribution to understanding the enzymatic mechanism. *Journal of Inorganic Biochemistry*, 98(5):862–877, May 2004. WOS:000221678300024.

[194] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of Physical Chemistry*, 98(45):11623–11627, 1994.

[195] Martin Tillmann Stiebritz and Markus Reiher. Hydrogenases and oxygen. *Chemical Science*, 3(6):1739–1751, 2012. WOS:000304365000001.

[196] Ludovico Sutto and Francesco Luigi Gervasio. Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. *Proceedings of the National Academy of Sciences*, 110(26):10616–10621, June 2013.

[197] 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, June 2009. WOS:000266929800003.

[198] Vitor H. Teixeira, Claudio M. Soares, and Antonio M. Baptista. Proton pathways in a [NiFe]-hydrogenase: A theoretical study. *Proteins-Structure Function and Bioinformatics*, 70(3):1010–1022, February 2008. WOS:000252836300035.

[199] Matteo Tiberti, Gaetano Invernizzi, Matteo Lambrughi, Yuval Inbar, Gideon Schreiber, and Elena Papaleo. Pyinteraph: A framework for the analysis of interaction networks in structural ensembles of proteins. *Journal of Chemical Information and Modeling*, 54(5):1537–1551, 2014. PMID: 24702124.

[200] B. L. Vallee and R. J. Williams. Metalloenzymes: the entatic nature of their active sites. *Proceedings of the National Academy of Sciences*, 59(2):498–505, February 1968.

[201] Henry van den Bedem, Gira Bhabha, Kun Yang, Peter E. Wright, and James S. Fraser. Automated identification of functional dynamic networks from x-ray crystallography. *Nature methods*, 10(9):896–902, September 2013.

[202] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, 31(4):671–690, March 2010.

[203] Michele Vendruscolo. Determination of conformationally heterogeneous states of proteins. *Current Opinion in Structural Biology*, 17(1):15–20, February 2007.

[204] P. M. Vignais and A. Colbeau. Molecular biology of microbial hydrogenases. *Current issues in molecular biology*, 6:159–188, 2004.

[205] Paulette M. Vignais and Bernard Billoud. Occurrence, classification, and biological function of hydrogenases: An overview. *Chemical Reviews*, 107(10):4206–4272, October 2007. WOS:000249839900012.

[206] Janice Villali and Dorothee Kern. Choreographing an enzyme's dance. *Current opinion in chemical biology*, 14(5):636–643, October 2010.

[207] Saraswathi Vishveshwara, Amit Ghosh, and Priti Hansia. Intra and inter-molecular communications through protein structure network. *Current Protein & Peptide Science*, 10(2):146–160, April 2009.

[208] A. VOLBEDA, MH CHARON, C. PIRAS, EC HATCHIKIAN, M. FREY, and JC FONTECILLACAMPS. Crystal structure of the nickel-iron hydrogenase from *Desulfovibrio gigas*. *Nature*, 373(6515):580–587, February 1995. WOS:A1995QG99700044.

[209] A. Volbeda, E. Garcin, C. Piras, A. L. deLacey, V. M. Fernandez, E. C. Hatchikian, M. Frey, and J. C. FontecillaCamps. Structure of the [NiFe] hydrogenase active site: Evidence for biologically uncommon fe ligands. *Journal of the American Chemical Society*, 118(51):12989–12996, December 1996. WOS:A1996WA82700015.

[210] A. Volbeda, L. Martin, C. Cavazza, M. Matho, B. W. Faber, W. Roseboom, S. P. J. Albracht, E. Garcin, M. Rousset, and J. C. Fontecilla-Camps. Structural differences between the ready and unready oxidized states of [NiFe] hydrogenases. *Journal of Biological Inorganic Chemistry*, 10(3):239–249, May 2005. WOS:000229092900004.

[211] CP WANG, R. FRANCO, JJG MOURA, I. MOURA, and EP DAY. The nickel site in active *Desulfovibrio baculatus* [nifese] hydrogenase is diamagnetic. *Journal of Biological Chemistry*, 267(11):7378–7380, April 1992. WOS:A1992HN48500027.

[212] H. X. Wang, C. Y. Ralston, D. S. Patil, R. M. Jones, W. Gu, M. Verhagen, M. Adams, P. Ge, C. Riordan, C. A. Marganian, P. Mascharak, J. Kovacs, C. G. Miller, T. J. Collins, S. Brooker, P. D. Croucher, K. Wang, E. I. Stiefel, and S. P. Cramer. Nickel l-edge soft x-ray spectroscopy of nickel-iron hydrogenases and model compounds - evidence for high-spin nickel(II) in the active enzyme. *Journal of the American Chemical Society*, 122(43):10544–10552, November 2000. WOS:000165205000006.

[213] Mei Wang, Lin Chen, and Licheng Sun. Recent progress in electrochemical hydrogen production with earth-abundant metal complexes as catalysts. *Energy & Environmental Science*, 5(5):6763–6778, May 2012. WOS:000303251500011.

[214] Wei-Ning Wang. Comparison of CO_2 photoreduction systems: a review. *Aerosol and Air Quality Research*, 2014.

[215] Christopher D. Wassman, Roberta Baronio, Ozlem Demir, Brad D. Wallentine, Chiung-Kuang Chen, Linda V. Hall, Faezeh Salehi, Da-Wei Lin, Benjamin P. Chung, G. Wesley Hatfield, A. Richard Chamberlin, Hartmut Luecke, Richard H. Lathrop, Peter Kaiser, and Rommie E. Amaro. Computational identification of a transiently open l1/s3 pocket for reactivation of mutant p53. *Nature Communications*, 4:1407, January 2013.

[216] Katharina Weber, Tobias Kraemer, Hannah S. Shafaat, Thomas Weyhermueller, Eckhard Bill, Maurice van Gastel, Frank Neese, and Wolfgang Lubitz. A functional [NiFe]-hydrogenase model compound that undergoes biologically relevant reversible thiolate protonation. *Journal of the American Chemical Society*, 134(51):20745–20755, December 2012. WOS:000313154200030.

[217] H Weinstein and E L Mehler. Ca2-binding and structural dynamics in the functions of calmodulin. *Annual Review of Physiology*, 56(1):213–236, 1994.

[218] Nils Widderich, Marco Pittelkow, Astrid Höppner, Daniel Mulnaes, Wolfgang Buckel, Holger Gohlke, Sander H. J. Smits, and Erhard Bremer. Molecular dynamics simulations and structure-guided mutagenesis provide insight into the architecture of the catalytic core of the ectoine hydroxylase. *Journal of Molecular Biology*, 426(3):586–600, February 2014.

[219] Christopher J. Woods, Maturos Malaisree, Benjamin Long, Simon McIntosh-Smith, and Adrian J. Mulholland. Analysis and assay of oseltamivir-resistant mutants of influenza neuraminidase via direct observation of drug unbinding and rebinding in simulation. *Biochemistry*, 52(45):8150–8164, November 2013.

[220] Hong Wu and Michael B. Hall. Density functional theory on the larger active site models for [NiFe] hydrogenases: Two-state reactivity? *Comptes Rendus Chimie*, 11(8):790–804, August 2008. WOS:000258393100002.

[221] Jenny Y. Yang, R. Morris Bullock, Wendy J. Shaw, Brendan Twamley, Kendra Fraze, M. Rakowski DuBois, and Daniel L. DuBois. Mechanistic insights into catalytic h-2 oxidation by ni complexes containing a diphosphine ligand with a positioned amine base. *Journal of the American Chemical Society*, 131(16):5935–5945, April 2009. WOS:000265460200043.

[222] Jian Yu. Bio-based products from solar energy and carbon dioxide. *Trends in Biotechnology*, 32(1):5–10, January 2014.

[223] G. Zampella, M. Bruschi, P. Fantucci, and L. De Gioia. DFT investigation of h-2 activation by [m(NHPnPr(3))(' s3 ')] (m = ni, pd). insight into key factors relevant to the design of hydrogenase functional models. *Journal of the American Chemical Society*, 127(38):13180–13189, September 2005. WOS:000232170800045.

[224] Yan Zhao and Donald G. Truhlar. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *Journal of Chemical Physics*, 125(19):194101, November 2006. WOS:000242181800003.