Universitat Autònoma de Barcelona

Computational design of Artificial Metalloenzymes

Lur Alonso Cotchico

Computational design of Artificial Metalloenzymes

Lur Alonso Cotchico

A Dissertation

Presented to the Faculty

of Autonomous University of Barcelona

in Candidacy for the Degree

of Doctor of Philosophy

Recommended for Acceptance

by the Doctoral Programme of

Biotechnology

Advisers: Professors Jean-Didier Maréchal and Agustí Lledós

Tutor: Jordi Joan Cairó

June 2018

Computational design of Artificial Metalloenzymes

Lur Alonso Cotchico

Recommended for Acceptance By Advisers

Jean-Didier Maréchal Agustí Lledós

June 2018

”Si no le temes a Dios,

témele a los metales.”

GABRIEL GARCÍA MÁRQUEZ

100 años de soledad.

Official acknowledgements

I would like to thank the *European Social Fund* and *Generalitat de Catalunya* for their funding support through their FI scholarships, which has made this Ph.D. thesis possible. The CTQ2017-87889-P financial support of the spanish *Ministerio de Economía y Competitividad* (MINECO) project is also acknowledged.

Moreover, I acknowledge Prof. Dr. Gerard Roelfes for his financial support and for allowing my stay at his laboratory during the last months of this Ph.D.

Acknowledgements

I have many reasons to be thankful and to think that I am lucky:

Be in the correct moment at the correct place when attending a lesson by Prof. Jean-Didier Maréchal, at the Master Degree, about theoretical chemistry. Feel how his passion for the field is directly transmitted to me. Achieve he finds my mail between a thousand of mails in his impossible mail list asking for practices at the theoretical group of chemistry ("I'm a biologist but...").

Receive the blind trust of my directors Prof. Jean-Didier Maréchal[1] and Prof. Agusti Lledós[2] from the very beginning to become part of the Transmet and Insilichem groups. Today I can say that this chess movement was one of the wisest of my academic life. Loss the count of the days Agusti spent next to me teaching from the base everything related with quantum calculations and all the concepts around. Feel at that time how my initial enthusiasm grew up exponentially. Receive academic therapy and hundreds of personal conversations from Jean-Didier during these years.

Share this experience next to a kind of people commonly known as theorists: Manu, Laia, Eli, Victor, Almudena, Pablo, Giuseppe, Jaime and Jose. I'm so proud to see how the time makes us to improve in all senses. Also Pietro, who guided me through the learning of Molecular Dynamics simulations (again learning from a best one). My friend Giuseppe, our conversations, discussions and your routine until 9 pm has greatly facilitate this hard pre-deposit period.

Enjoy students any supervisor would desire to have: Eric and Lorea. Lorea, I am not sure if you learned more from me than I did from you.

[1]Our academic father. A person with the ability to make students to grow up ten years in the course of three.

[2]What happens if you mix an internationally recognized 60 year old man with the most rocking and humble person of the world? That Agusti passes running with his "Machete" t-shirt (this by default, a Santa Claus one if is Christmas).

Share practices, lunch time, post-lunch time, parties and epic moments also with all the people around, involving noteworthy discoveries like Gregori, Egil, Javi and Sergi.

Have a family far from my home town, Jaime and Sonia. Share everything, feel supported always and judged never. You both are part of me and you make me feel lucky.

Be involved in a project about Artificial Metalloenzymes. That the man behind that project is Prof. Gerard Roelfes (The University of Groningen). Receive his trusting and his good words always from the distance to maintain my motivation level on the top. I couldn't be more thankful. Accept me as a proper member of his group. Bring with him an army of good Ph. D. students with which the collaborations couldn't be better: Ivana, Lara, Cora, Yans...

Have the opportunity to also collaborate with other powerful experimental groups: Prof. Tomas Ward (The University of Basel) and Prof. Eugenio Vázquez (The University of Santiago de Compostela). Our work together has been (and continue being) essential for the validation of the implemented computational frameworks and I am so thankful for your trust on my work.

Cross the path of people that makes you remember that the best part of the academic life is to know new people (See Fahmi and Sheila for further details).

Close this story travelling to Groningen. Learn that things can be much better than expected, even the weather. Feel at home from the first day thanks to Cora, Max, Nikola, Elvira, Misun, Kate, Mohammed, Simone, Friso, Sara, Sambi, Linda and Hemant. Thanks for a better impossible time in the city (to be continued...).

Have my family always close despite the distance. Everyday I am concious about how lucky I am to have you all.

I am not sure if there is still something remaining from the girl that once started the Ph. D. But in conclusion: I am thankful, I am lucky and I am happy.

Abstract

Enzyme design is the scientific field that aims at discovering and/or optimizing biomolecules to reach new-to-Nature reactions. It is an area in wild expansion and constitutes one of the cornerstones of the transition of chemical practices towards greener alternatives.

An elegant way to construct novel biocatalysts is through the embedding of organometallic cofactors into biological scaffolds, leading to the so-called Artificial Metalloenzymes (ArMs). Conceptually mimicking the naturally occurring hemoenzymes, those hybrids bridge the catalytic versatility of the organometallic compound with the substrate and spatial specificity of the biological host. The design of ArMs has spread increasingly during the last two decades taking clear advantage of the major expansion of structural biology and the maturity of organometallic catalysis. In this area, many molecular variables need to be controlled (i.e. the location of the artificial cofactor, the dynamics of the coordination sphere of the metal, the impact of the second coordination sphere provided by the biological scaffold, etc.). Molecular modelling aims to help designers to provide with structural information that could serve for constructing optimum biocatalysts. However, despite the increasing improvement of the computation performance and the exponential development of new simulation techniques, the complexity of dealing with transition metal including systems has promoted modellers not to explore the ArM constructs.

The InSiliChem group, in which this Ph.D. has been performed, has focused on developing a specific framework for the study and design of ArMs. In particular, this has been based on the development of *in silico* multiscale strategies including standard computational methods. Prior to this work, these protocols mainly incorporated quantum based (Quantum Mechanics and the hybrid Quantum Mechanics/Molecular Mechanics), structural statistics (structural bioinformatics) and force-field based approaches (Protein-ligand docking). However, two limitations existed: First, the pro-

tocol did not introduce extensive sampling of the pre-catalytic space (meaning the ensemble of geometries of the system susceptible for the reaction to proceed). Second, regarding the applicative side, it was used to elucidate the mechanism of previously characterized ArMs but was never tested for *de novo* enzyme discovery and optimization, the true goal of enzyme design.

This Ph.D. was aimed at increasing the potentiality of our computational platform for ArM design by 1) including classical Molecular Dynamics simulations into the integrative computational framework and 2) testing the validity of the methodology for the design of real case ArMs.

The results obtained could be summarize as follows:

1. The catalytic mechanism of two novel ArMs were decoded using the updated computational pipeline. These were a copper-Phenanthroline containing hydratase based on the Lactococcus Multidrug Resistance Regulator (LmrR) and a variety of novel mutants based on Streptavidine-Noyori complexes for cyclic imine reduction reaction. From the technical point of view, their study leaded to two major challenges: the optimization of the workflow for the parametrization of the metallic systems inside the AMBER force field and the incorporation of GPUs accelerated sampling. The study revealed the importance of the contribution of the MD simulations to decode the catalytic mechanism of these ArMs and to assess the impact of second sphere mutations on the catalytic tendencies. (Chapter 4)

2. Using the same approach, calculations were carried out for the *in silico* design of hydratases, but in this case based on the inclusion of a novel unnatural amino acid. This was first applied to the LmrR scaffold, for which mutants suggested via computation for optimum enantiomeric excess (ee) were then experimentally assessed with success. Next, based on the experience obtained, we expanded the *de novo* exercise towards the identification of peptidic scaffolds able to support

the inclusion of the same unnatural amino acid to become efficient hydratases. (Chapter 5)

3. The final part of the work focused on deciphering molecular variables that our previous studies showed to be far more complex than expected. This was the impact of the active site configuration to define the catalytic activity of the ArMs. In particular, we decoded the rearrangement of a variety of LmrR-heme complex for the cyclopropanation reaction to proceed. From this study it clearly appeared that the flexibility of the receptor is key for the porphyrin based ArMs to reach their catalytic activity. To further assess the importance of this molecular variable, we expanded this work to the study of distinct naturally occurring heme binding proteins. (Chapter 6)

Overall, this Ph.D. represents a step forward on the methodological development of the computer based enzyme design. Furthermore, it sheds light on how transition metal compounds could cooperate with biological scaffolds at the molecular level with the focus on the *de novo* design of new biocatalysts.

List of Abbreviations

- **AIMD**: *ab initio* Molecular Dynamics
- **ArM**: Artificial Metalloenzyme
- **BpyA**: Ala-2,2'-bipyridine
- **ee**: enantiomeric excess
- **LCP**: Latex Cleavage Protein
- **LmrR**: Lactococcal multidrug resistance Regulator
- **MM**: Molecular Mechanics
- **MD**: Molecular Dynamics
- **QM**: Quantum Mechanics
- **QM/MM**: Quantum Mechanics/Molecular Mechanics
- **PDB**: Protein Data Bank
- **Phen**: Phenanthroline
- **RESP**: Restrained Electrostatic Potential
- **SMD**: Solvation Model based on Density
- **Sav**: Streptavidin
- **TS**: Transition State
- **UAA**: Unnatural Amino Acids

Contents

Chapter 1

Introduction

1.1 Taking advantage of Nature's benefits

Since prehistory humans have been conscious about the strengths of Nature machinery and have exploited it to achieve their own purposes, ranging from the elaboration of alimentary products like wine, beer and cheese to its application on areas such as pharmacology and industry. Taking advantage of biological catalysts or *enzymes* to carry out chemical transformations is known as *biocatalysis*. [1]

At the end of 19th century and the beginning of 20th biocatalysis emerged as a new technology with scientific potential. With the development of well-established biotechnological approaches, such as protein engineering, enzyme immobilization, managing living cells or bioinformatics, [1–4] it has been implemented in many successful chemical processes, being the most famous cases related to C-O/C-N reductases and oxydases, acyl transferases and glycosidases. [5]

Biocatalysis not only represents a powerful way for controlling chemical reactions in terms of catalytic specificity and enantiosectivity, [2–5] it is also paves the way for the transition of chemical industries towards greener practices. [6] However, the high specificity that characterizes the naturally occurring enzymes to recognize a narrow

substrate scope drastically reduces their field application. Tricking Nature's enzymes to create new ones that meet real industrial benefits is a raising area of research lying at the midpoint between biology, chemistry and physics. This is the field of *enzyme design*.

1.2 Enzyme design approaches

Tailoring new efficient enzymes usually encompasses two steps: **discovery**, which is related to the detection of a new receptor or substrate after a rational or random search process, and **optimization**, in which the properties of the new enzyme like yield, selectivity or specificity are iteratively improved to achieve the desired objective. Depending on the goals of the researcher and the raw material available at the initial stage of the study, the design of an enzyme with a novel activity could require major efforts in the first, second or both steps.

1.2.1 The enzyme discovery stage

Depending on the strategies employed for achieving new enzymes, these can be grouped as:

***De novo* enzymes**

This category includes: on one hand, those enzymes generated "from scratch", i.e. their whole structure is absent from Nature; on the other hand, those generated from existing folds but provided with absolutely different functions.

The ideal situation to generate *de novo* enzymes would be getting "from scratch" non-natural scaffolds able to satisfy pre-defined features related to a determined catalytic activity, substrate specificity or enantio- and regioselectivity. However, achieving the most stable geometry without any structural information, would require the

massive exploration of the conformational and chemical spaces as well as an extensive mutational analysis(Figure 1.1 a), a context which nowadays is out of range. More accessible ways to obtain usefull tertiary structures results from implementing complex algebraic equations to achieve feasible backbone geometries based on "sequence-independent" design principles, [7–15] followed by a process of sequence sampling over the pre-stated backbone structures (Figure 1.1 b); or, alternatively, joining pre-existing structural "building-blocks" from functional proteins (Figure 1.1 c). [16–18] These strategies result more achievable since designs are based on well-defined structural patterns. Some successful approaches which follow the "building-blocks" philosophy are the *de novo* enzymes based on helical bundles [8,9,11,12,19–23] and the pharmacologically active disulfide-crosslinked peptides. [23, 24]

Also a variety of computational tools exits which help to predict *de novo* folds from available primary sequences based on tertiary structures of phylogenetically related proteins. Probably the best recognised are the *Homology Modelling* (HM) [25] and *Rosetta* [26, 27] approaches. The former is quite useful when existing 3D structures available at the *Protein Data Bank* (PDB)[1] [28] with a sequence similarity of a minimum of 25-30% to the target sequence. The later uses algorithms based on pre-existing structural patterns to model *de novo* folds and has demonstrated a very good performance when predicting the folding of sequences up to 125 amino acids. [27]

An alternative and more accessible trend for the *de novo* enzyme design consists in adapting natural non-catalytic scaffolds to provide them with a determined catalytic activity (Figure 1.1 d). The changes both can consist in performing random point changes along the protein scaffold via *directed evolution* techniques, [29–34] or can be rational and focused on certain areas of the protein. The main limitation of altering

[1]The Protein Data Bank is probably one of the most important contributions to structural biology due to the exponential increase in the amount of registered structures of proteins, nucleic acids and complex assemblies since its foundation in 1971. Consisting initially of only 7 structures, nowadays contains structural information of around 137,000 proteins and more than 24,600 ligands.

the native protein scaffold is that cumulative modifications can lead to substantial structural changes [35] that, in the worst case scenario, can promote unfolding or aggregation.

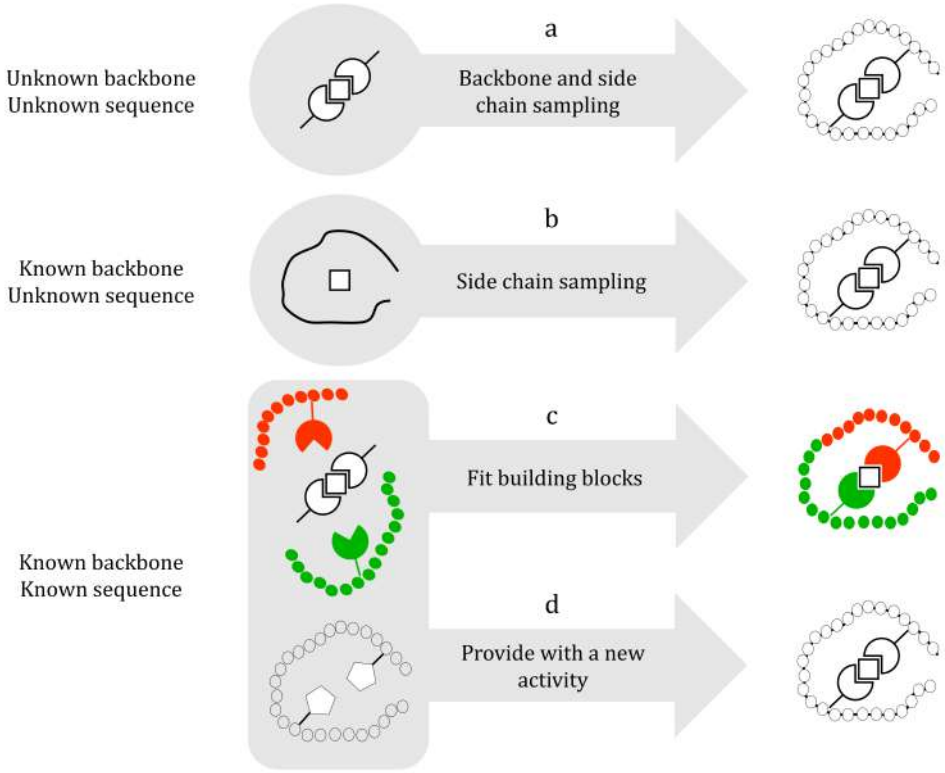


Figure 1.1: Strategies for the *de novo* enzyme design.

Re-designed enzymes

This group encompasses catalytically efficient enzymes that have been modified by performing biochemical changes on the scaffold, the cofactor (if any) or the substrates to achieve different chemical reactions or to improve properties such as yield, selectivity or substrate specificity. Here, promiscuos scaffolds appears promissing to modulate the catalytic outcomes since their versatility easies the process to drive their preference for the adequate catalytic pathway. [36] Depending on the biological level at which the system is manipulated, two main categories can be differentiated: *sequence-based* and *structure-based* enzyme re-design.

Sequence-based enzyme re-design The approaches comprised in this group combine *Multiple Sequence Alignments* (MSAs) with phylogenetic analysis to identify functional "islands" of conserved amino acids at the protein sequence and ancestral relationships between proteins. Then, these hot spots are submited to local amino acid variability to generate new designs. Some helpful computational tools which facilitate the data analysis stage are: the HotSpot Wizard server, [37] which combines structural information from homology models or 3D-structures with extensive sequence database to generate the mutations; [38] the 3DM database, [39] which, additionally, is able to update itself from literature with new structural and functional information; [40–42] the *Reconstructing Evolutionary Adaptive Paths* (REAP), [43] which is based on phylogenetic studies on ancestral proteins to identify key mutations lying at functional divergence points during evolution to generate functional enzyme libraries; and ProSAR, [44] a directed evolution methodology described by G.W. Huisman et. al for performing statistical analysis based on protein activity-sequence patterns to facilitate the identification of advantageous mutations.

Structure-based enzyme re-design The rapid increase of the amount of 3D-structures into the PDB together with the rapid advances on computational approaches (specially homology modelling tools) allows protein engineers to identify functional residues for performing more specific catalytic models. These strategies are generally based on generating libraries of mutants that are tested and filtered to identify the best candidates for specific purposes. Getting a successful candidate highly depends on the quality of the pre-stated library of mutants, being usually the most limiting step of the process. To ensure building good quality libraries, these strategies are usually initiated by computational inputs which help to guide the design of the variants and also allow scientists to achieve their proposals with less effort and in a much faster way. This has resulted as a powerful approach for modifying

the cofactor or substrate specificities by altering not only amino acids located at the active site [45–48] but also amino acids lying at further distances that through "cascade" effects ultimately promote functional benefits. [49–51]

As a noteworthy example, Reetz and coworkers stated the iterative use of *Iterative Saturation Mutagenesis* (ISM) and *Combinatorial Active-site Saturation Mutagenesis* (CAST) as a powerful alternative when only low or medium throughput screening approaches are available. [52] ISM directs mutations towards defined sets of up to three amino acids at the active site and then, for each set, all possible variants are screened via CAST. Only the improved mutants are iteratively submited to new rounds of ISM plus CAST allowing them to evolve towards the desired direction. A variety of successful designs exists that have stated their bases on this methodology supporting its high efficiency and versatility. [49, 52–56] A drastically different strategy is known as *chimeragenesis*, which focuses on the combination of structural moieties from different proteins or peptides for the construction of functional chimeras. A very effective computational framework for predicting feasible recombination of interchangeable fragments with minimal structural disturbance is SCHEMA, [57] which has been successfully applied to systems such as cytochrome P450, [58] β-lactamases [59] and cellobiohydrolases. [60, 61]

Artificial Metalloenzymes

At the middleground of the above categories lie the *Artificial Metalloenzymes* (ArMs), which result from including a catalytically active organometallic cofactor into the context of an asymmetric biological scaffold leading to non-natural reactions. This thesis has been focused on the study and the development of this growing area and thus, separate sections will be devoted later. Especial interest will be placed on emphasizing the latest developments in the field, regarding its application to organic chemistry and industry and the available computational tools that aid with the challenging task

that their design and optimization mean. A very recent book summarizes the best achievements in this area. [62]

1.2.2 Optimization strategies

The first generation of a non-natural enzyme may not have all the desired catalytic performance. Usually, further optimization steps are required to improve features such as yield, stability or substrate specificity, between others. For that purpose, two main strategies exist:

Rational re-design

The system is consciously optimized by performing rational changes in the close vicinity of the point where catalysis take place. [63] It is purely based on structural criteria and is carried out both via mutations or using non-canonical amino acids. [64, 65] As the number of newly generated variants is limited, it is a cheap approach but, on the counterpart, it requires initial structural and (bio)chemical knowledge about the system under study, something that is not always accessible.

Directed evolution

They are based on performing changes via a random and massive mutagenesis. The resulting library of variants is then subjected to high-throughput screening based on a selective pressure, in this case the desired reaction, and the more efficient ones are selected. This first generation of mutants is released to a new random process of mutagenesis, leading to the second generation of variants. The mutagenesis and screening processes are iteratively repeated until highly efficient scaffolds are obtained (Figure 1.2). [66–68] It has been demonstrated that the number of iterative cycles dictates the efficiency of the final candidates. [68] This strategy has been extensively used and has been proved to be highly successful to drive the enantioselective profiles of enzymes

towards a desired direction. [69] Since it is a "blind" process, its main advantage is that previous molecular information of the system is not required. In contrast, it is relatively time and resource consuming and obviously lacks any rationalization on the nature of the molecular origin of the enzyme improvement.

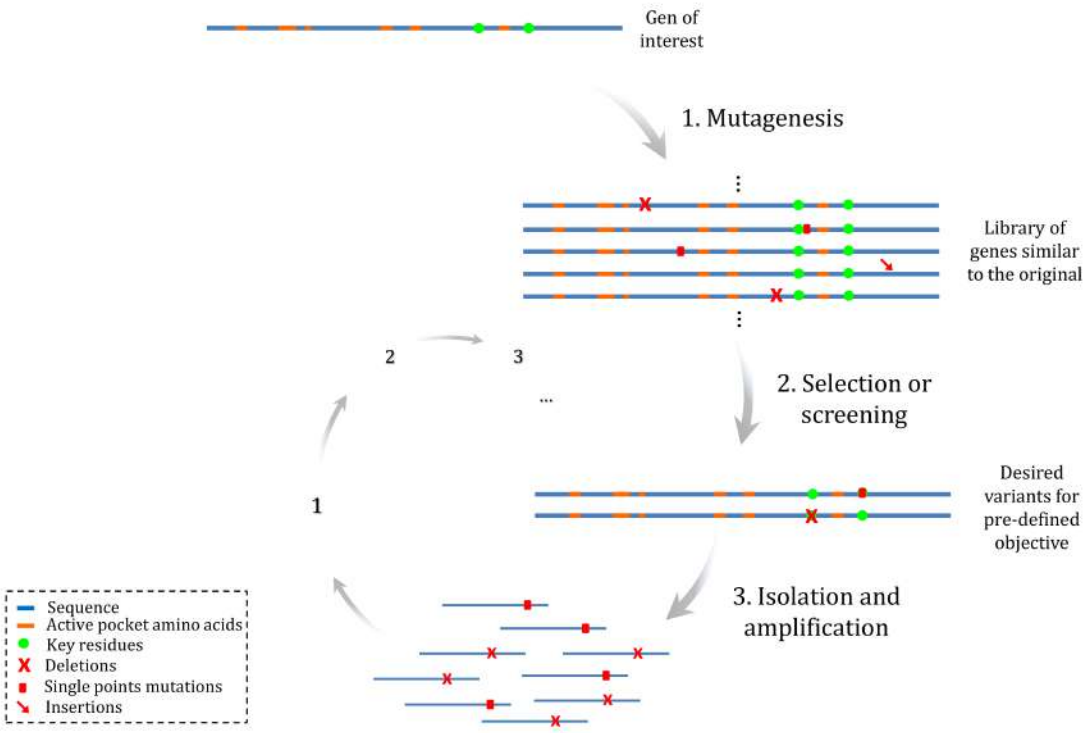


Figure 1.2: Directed evolution strategy.

During the last decades directed evolution approaches have transformed the area of protein engineering, being one of the main driving forces behind the understanding of protein structure and function. This has allowed large mutational libraries to evolve towards more focused functional ones, expanding the scope to the so called semi-rational designs. [70]

Semi-rational approaches

Aiming to bridge the advantages of both rational and directed evolution strategies, some research groups focus on performing knowledge-based random changes. When

X-ray structures of an enzyme are available, it is possible to direct these variations only to functional amino acids located at the active site that consequently are far more likely to influence catalysis. This strategy encompasses several advantages such as reducing the cost of the process regarding a pure directed evolution process and enhancing the probability of detection of cumulative effects than can have a strong impact on the catalytic mechanism. [70] However, changes on both long range interactions and the structural effects generated by amino acids lying at further distances from the active site (known as "cascade effects") which can be beneficial for catalysis [50, 71, 72] are neglected.

1.3 Artificial Metalloenzymes

1.3.1 An overview

Regarding catalysis, transition metals are of the most versatile elements of the periodic table, due to the special features comprised in their proper nature: [73]

- They have a large *charge/radius ratio* and this can vary depending on the coordination environment, acquiring negative, neutral or positive charge.
- They can interact with both *organic* and *inorganic ligands* and, depending on the *oxidation state*, the coordination number and geometries can vary giving rise to a large variety of shapes and complexes with interesting magnetic and electronic properties.
- The *Lewis acid* character of metals provides them with high electron affinities and promotes polarization of the ligands, easing chemical processes such as the hydration reaction.

- They have a *partially filled d-shell* that allows them to perform one electron oxidation and reduction reactions by interconversion between the different oxidation states

Nature has taken advantage of these special features by including transition metals into the context of biological scaffolds to provide them with structural benefits and/or allowing them to become efficient catalysts. In fact, they are present in one-third of the known proteins, [74, 75] in which are included the so-called metalloenzymes. [76] These are essential for the correct performance of many biological processes such as the transmission of nerve impulses and biological reactions like those involved in oxygen transport, energy transfer or photosynthesis. Additionally, the embedding of the metal into a biological cavity provides a protective environment which controls its oxidation state and reactivity, fact which results essential in life systems to prevent toxicity. For example, chromium (III) is required in glucose metabolism [77] but chromium (VI) results toxic and carcinogenic. Metalloenzymes have evolved over centuries and have been naturally selected to become the most efficient constructs for a specific task. Their activity is driven by the nature of the *first* and the *second coordination spheres* of the metal. The former is constituted by the nearby area of coordination to the metal and is the responsible of the most of the catalytic properties. The later corresponds to the asymmetric environment provided by the biological scaffold and modulates the catalysis by defining selective profiles and substrate specificity. This architecture makes them to be highly selective and to recognize a restrained substrate scope. [4] Although a variety of natural enzymes with interesting applicability exists, [5] that fact limits their use for industrial purposes.

Artificial Metalloenzymes (ArMs) arise as an elegant strategy to merge the versatility of transition metals with the selectivity and specificity of enzymatic catalysis. [78] This is achieved by including well-known organometallic cofactors or non-canonical aminoacids able to bind metals into the context of biological scaffolds (Fig-

ure 1.3). [79–83] This promotes the reorientation of the natural activity of the macromolecule to become novel biocatalyst. The successful application of the concept has boosted the interest in the field and the number of new ArMs has strongly increased over the last decade. The generation of new ArMs is still a continuing growing area driven by the design and identification of new ligand-host partners able to perform specific functions absent from Nature.

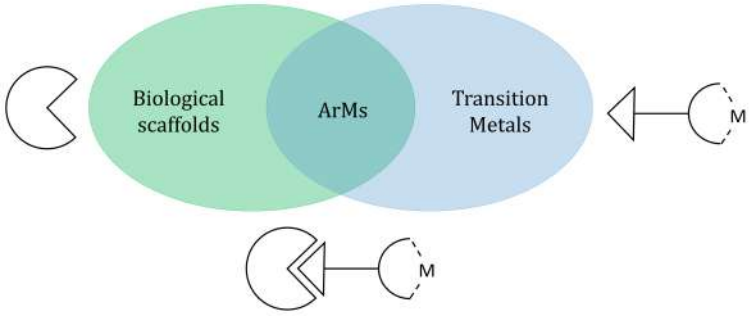


Figure 1.3: Graphic representation of the artificial metalloenzyeme concept .

1.3.2 Artificial Metalloenzymes design

Tailoring proteins for reaching non-natural activity is one of the major challenges for protein engineering and structural biology. For this purpose, the combination of chemical and structural knowledge as well as the understanding of the catalytic mechanism of novel enzymes results essential. [84]

Following the two steps process for the design of non-natural enzymes exposed previously (discovery and optimization), the first step in ArMs design at the **discovery** stage is the identification of an initial convenient partnership between the homogenous catalyst and the corresponding biomolecule.

The selection of the homogenous catalyst depends obviously on the type of catalysis that is targeted. This takes into account specific features like the nature of both the metal (if it performs redox or non-redox-reactions, if is catalytic or just assist catalysis, or if carries out electrostatic stabilization) [85] and the ligand it is bound

to (mono- or poly-dentate species such as amines, pyridines, tridentate Schiff bases or phosphines). One or a set of synthetic catalysts can be selected as part of the starting material to approach the final design.

Selecting a feasible scaffold (usually a protein nowadays) for experimental practices requires taking into account some fundamental features such as its good thermal and structural stability and its easiness of expression. Additionally, as the protein-cofactor complex has not been naturally optimized and the cofactor may differ a lot from the natural ligand, the host must be flexible and stable enough both to adapt its structure and to support mutations during the optimization stage.

The major challenge relies on getting a stable enzyme-cofactor-substrate triad able to behave in synergy for an efficient catalysis. [86] To facilitate this process, designs are usually based on previously solved X-ray structure with well-identified anchoring cavities. In the last years, a variety of successful biological scaffolds that own these qualities have been widely used for the design of ArMs [87] like Streptavidin (Sav), [88] Bovine Serum Albumin (BSA), [89–91] cytochrome P450, [92] Neocarcinostatin (NCS) [93] or the Lactoccocal Multidrug Resistence Regulator (LmrR). [79,94,95] The use of DNA strands as scaffold has also led to successful designs for enantioselective catalysis. [96–100]

To include the artificial cofactor into the protein scaffold different strategies can be used (Figure 1.4). This is referred to *supramolecular anchoring* if the embedding of the artificial cofactor into the host is driven by molecular recognition affinity via non-covalent interactions [36, 36, 89, 90, 101]; to *covalent anchoring* if the organometallic cofactor is directly linked to the protein scaffold by a covalent bond; [79, 95, 102] or to *dative anchoring* if the enzyme is rich in coordinating residues (those containing S, O or N functional groups) able to bind directly the unsaturated metal centre. [91, 103] Alternatively, the artificial cofactor can be introduced into a protein cavity via a "Trojan horse" by taking advantage of the high affinity that the host protein has

for another non-covalently linked ligand, as is the case of Streptavidin/avidin-biotin based ArMs. [88, 104]

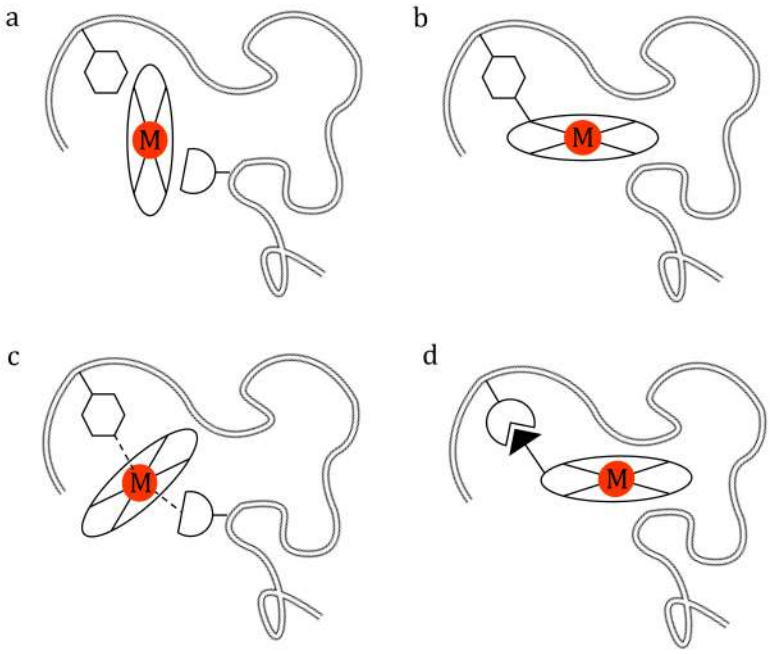


Figure 1.4: Anchoring strategies to firmly localize an organometallic cofactor into a biological host: a. supramolecular, b. covalent, c. dative anchoring and d. Trojan horse strategy.

Sometimes, the presence of stabilizing and/or reactive residues at the active site makes the initial design to be catalytically efficient. [87, 94] However, this is not the normal case and, usually, further optimization is required to reach specific chemical purposes such as improved yield or enantioselective profiles.

The **optimization** stage is usually focused on modifying the protein scaffold in the close surroundings of the catalytic area to generate a library of variants with optimized performance. [105, 106] The successful ArMs can constitute the base for additional (usually rational) re-design (Figure 1.5), which could involve a large variety of biological and/or chemical changes on the host-cofactor-substrate complex like:

- Change the *biological scaffold* for another which is able to satisfy pre-defined structural requirements, both for fixing the cofactor and properly allocating the substrates to perform catalysis of the same or improved quality. [36, 98]

- Try different *substrates* able to modulate the catalytic mechanism, the molecular recognition or the type of catalysis. [105]

- *Metal* and/or *cofactor* substitution into the same receptor to give raise to improved or to other new-to-nature reactivities. [94, 95, 107]

- Perform *rational mutagenesis* on the active site amino acids either to alter the chemistry of the system or to adapt the receptor to variations performed on the substrate or the cofactor. [95, 105]

- Modify the location of the catalyst into the protein scaffold, providing a different chemical environment as well as a new vacant place for the accommodation of the substrates (no reported cases yet).

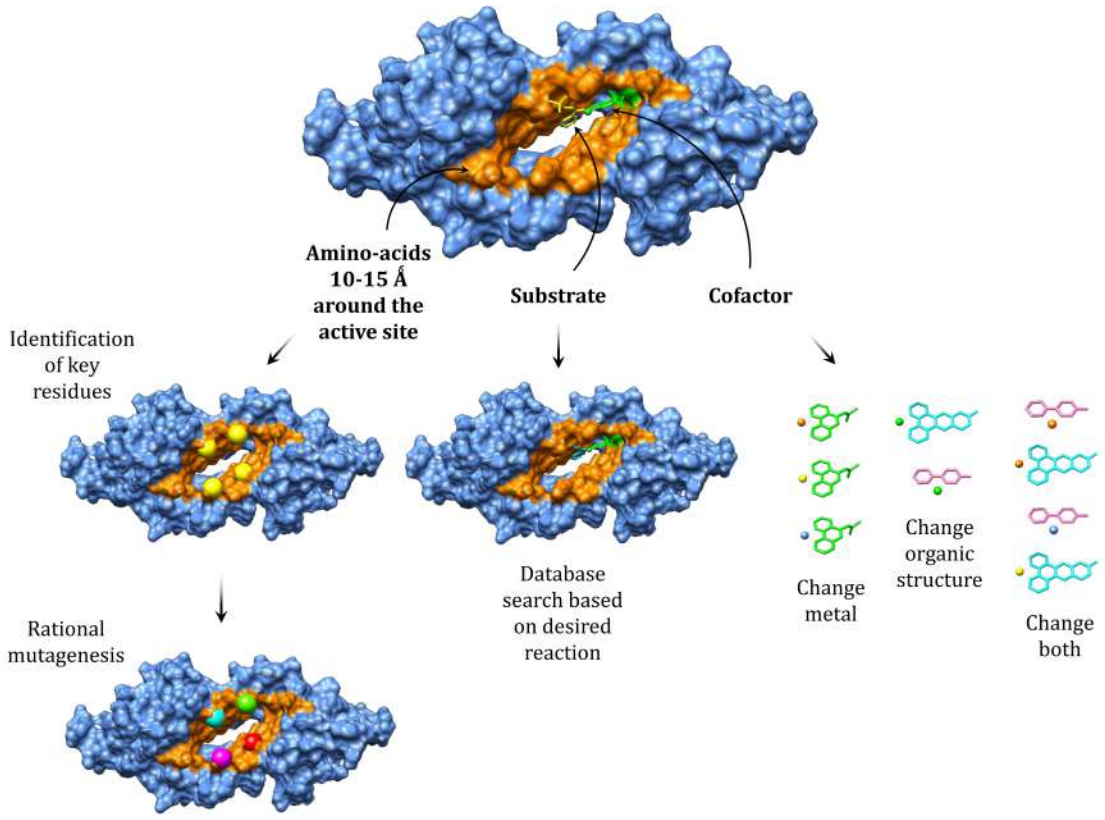


Figure 1.5: Rational ArM re-design at the optimization stage.

1.3.3 The revival of the field

The first asymmetric synthesis catalysed by a transition metal aided by a protein was reported in 1956 by S. Akabory and coworkers, [108] which performed the asymmetric hydrogenation of dehydroamino acid derivatives based on silk fibroin impregnated with a palladium catalyst. Later, during the 60's and 70's, some studies on natural metalloenzymes were done based on metal substitution, some of which experienced substantial changes in either the activity [109, 110] or selectivity [111] of the enzyme. Even worth mentioning as they comprise the first attempts in enzyme design, all these constructs lie out of the concept of ArMs which, by definition, considers an organometallic cofactor into the non-native context of a biological scaffold. Based on this, in 1978 Wilson and Whitesides performed the first series of ArMs by embedding

an achiral diphosphinerhodium(I) catalyst into a variety of protein scaffolds to get the asymmetric hydrogenation of the α-acetamidoacrylic acid. [112] During the following decades the field failed to attract much attention and only a few examples of enzymatic enantioselective catalysis were reported, mainly based on the bovine serum albumin (BSA) protein. [91, 113]

It has not been until this century that the developments on structural biology and protein engineering have promoted a major expansion of the field. [87] The successful outcomes have caught the interest of scientist community and have boosted the revival of the area, promoting the exponential increase of the successful designs for reaching interesting chemical reactions such as C-C bond formation, C-H activation, olefin metathesis, oxygen insertion, hydration and reduction processes, with impressive enantioselective levels in many cases. For a more extense information the reader can refer to the variety of nice recent reports that exist comprising the most successful advances on the ArMs design during the last decades. [62, 73, 87, 114, 115]

1.4 *In silico* design strategies

The computational framework for studying a biocatalyst must account for events occurring at different length and time scales that, in conjunction, describe a certain process. This usually requires accounting for computational approaches based on different levels of theory that are combined leading to *multiscale strategies*. These can be of very different nature depending on: the stage of the design the researcher is (discovery or optimization), the amount of available information when initiating a case of study and the final goals (searching conformations, identify new scaffolds, substrates or cofactors, analysis of the reaction mechanism, modifying the selectivity, etc). [116] Next sections comprise the most relevant applications of the *in silico* approaches to the enzyme design field, grouped by the different nature of the designs (*de*

novo or re-designs). Last, a separated section especially focused on ArMs designed via computation will be devoted to close the Chapter.

1.4.1 *De novo* designs

A remarkable *de novo* design was performed in 2001 by Mayo and coworkers, when converted the non-catalytic *E. coli* thioredoxin into an effective esterase. [117] Via computation, they were able to identify catalytic positions and stabilizing mutations which drastically increased the rate of the p-nitrophenyl acetate hydrolysis reaction. Their strategy was then expanded by Baker and coworkers for a large variety of successful *de novo* designs able to perform reactions absent from Nature, including Kemp eliminases, [118] Diels-Alderases [119] and Retro-Aldolases [120]. These were modelled based on their own stated protocol which comprised, first, the design of a library of catalytic models known as "theozymes", followed by their inclusion into protein scaffolds able to provide with the adequate environment with the RossetaMatch algorithm. [121] The new designs were then optimized via rational changes and directed evolution techniques for enhancing catalytic activity. [118, 122–125] The same group continued extending their strategy both to find potential protein scaffolds able to catalyse the Morita-Baylis-Hillman (MBH) reaction [126] and to construct a variety of artificial sterases. [127, 128]

Separately, Ruscio and coworkers supported the relevance of incorporating Molecular Dynamics (MD) simulations into the design protocols to get adequate optimized models of an alternate retroaldolase enzyme. [129, 130] More interestingly, Lu and coworkers combined homology modelling approaches with MD simulations to reproduce the active site of the nitric oxide reductase (NOR) -an important protein in signal transduction pathways for which no structural information was available- into the myoglobin scaffold for the design of a catalytically active NOR model [131] that was used as starting point by the same group for further re-design. [132]

Between the most popular *in silico* strategies for the *de novo* enzyme design are those reported by Houk, Mayo and Backer, [127] which perform initial Quantum Mechanics (QM) calculations to identify the TS geometries and the chemostructural environment required for its stabilization. Then, a reduced model of the active site called "theozyme" [133] is generated, followed by the identification of natural hosts that, once optimized, are able to fit these TS geometries. This screening process is performed by search algorithms that help to fit the described theozyme with the active site geometries of well-known protein structures from the PDB database. [134] Programs such as RosettaMatch, [121] Dezymer, [135] ORBIT, [136] OptGraft, [137] PRODA_MATCH, [138] SABER [139] or ScaffoldSelection [140] help to perform this task.

1.4.2 Re-designed enzymes

Re-design protocols are usually used by protein engineers to enhance protein stability or catalytic activity [45] and have also led to a wide range of successful protein models. In this context, Protein-ligand docking technique has been demonstrated to be a powerful tool at initial stages of the designs followed by rational [141] or semi-rational [51] optimization. Proteins such as Cytochrome P450, [141] glucose 6-oxydase [142] and S-adenosylmethionine synthase (SAMS) [143] have resulted adequate hosts for re-designing based on docking approaches. This approach has been demonstrated to be highly versatile for the study of a large variety of systems when refined via MD simulations [144–147] and full QM or hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) approaches. [148].

QM/MM calculations have been widely implemented also as part of multi-stage protocols in recent years for the optimization of enzymatic systems. [149,150] Remarkable examples are the optimization of Kemp eliminases by Houk and coworkers by combining QM, QM/MM and MD simulations; [151] or the engineering of the Lacasse

enzyme carried out by Victor Guallar and coworkers by combining MD simulations and QM/MM approaches. [152]

Other successful protocols are based on the *in silico* generation of initial libraries of mutants that are then analysed and filtered based on chemical interests. More specifically, Codexis program was used by Houk and coworkers to generate variants of a ketoreductase (KRED) that were then analysed via MD simulations. [153] In the same way, the very interesting Framework for Rapid Enzyme Stabilization by Computational libraries (FRESCO) workflow reported by Wijma et al. based on combining structure prediction programs with MD based screening steps was used to generate functional variants of a limonene epoxide hydrolase. [154] Alternative computational tools which help to overcome the challenges that protein re-designing encompasses are: OptZyme, which aims at reducing the TS barriers occurring at the active site by rational redesign and was used to improve the catalytic activity of the E. coli α-glucuronidase (GUS); [155] the Dynamic Disulfide Discovery (DDD) algorithm, which attempts to predict the disulphide bonds that can be formed in a protein by analysing frames of the MD simulation and was the first step of the previously mentioned FRESCO workflow; [154] and the K* algorithm developed by Chen et al., that was used to successfully identify positons at the active site to switch the substrate specificity of the gramicidin S synthetase enzyme. [156]

As last relevant case, Murphy and coworkers demonstrated that re-designs do not need to be limited to one or to a reduced number of mutations, but they can be based on the replacement of entire fragments of the proteins. Taking advantage of the Rosetta Design algorithm [157] they performed the re-design of the human guanine deaminase (hGDA) through the substitution of entire loops achieving high specificity for ammelide substrate while decreasing the activity for guanine, which meant an important advance for suicide gene therapy. [158]

1.4.3 Artificial Metalloenzymes

Accounting for the insertion of homogeneous catalysts into a biological scaffold enhances the difficulty of the study due to the necessity of predicting the complementarity and interactions between three entities: the biomolecule, the cofactor and the substrates. Additionally, the most commonly used force-fields do not include parameters for the majority of organometallic interactions and these need to be properly parameterized. This task is hampered by the fact that, the binding could lead to modifications on the homogeneous catalyst related with changes on the metal first coordination sphere or on its electronic state, or to structural rearrangements on both the cofactor and/or the protein host. All these reasons promote that the majority of the *in silico* studied enzymes containing metals are related to systems composed by metals directly linked to the biomolecular scaffold (no ArMs by definition), and are based on modifying the close surroundings of the metal, the metal itself or on finding new coordinating spaces along the protein surface. [159–161]

Among the few reported examples of *in silico* aided ArMs desings is the case of study of the Nika protein by Ménage and coworkers. They were able to include and stabilize different iron containing cofactors via the protein-ligand docking technique. Further computational screening allowed to recognize substrates with best catalytic potential that were then experimentally supported. [162, 163] Separately, Morokuma and coworkers were the first in rationalizing the catalytic mechanism of an ArM by QM/MM approaches. [164] Based on the ONIOM scheme they were able to study de mechanism of the apo-Ferritin protein containing a $[Rh(norbornadiene)Cl]_2$ catalyst and elucidated the best cavity into the protein for the polymerization of the phenylacetylene substrate. Probably the most relevant integrative strategies found into the field are the ones performed by Houk and Baker. [33] They tried different multiscale protocols based on the combination of QM, full QM cluster models, QM/MM full size protein models and classical MD simulations. The inclusion of MD simulations into

the protocol elucidated substantial differences on the side chain and backbone dynamics in solution that appeared to be critical for the proper differentiation amongst the active and the non-active entities.

Our group is focused on overcoming the issues related with the study of ArMs and has reported a variety of studies based on multiscale strategies for their specific study, [86, 95, 105, 165–171] some of them included into this thesis. These comprise from ligand-protein docking based protocols to complex integrative strategies for the rationalization and optimization of different artificial systems and include at least two of the next steps: 1) the QM study of the catalytic mechanism for elucidating the TS geometries, 2) the docking of specific cofactor-substrates complexes into the protein scaffold, 3) the analysis of the dynamical behavior of the complex, 4) the evaluation of energy barriers via QM/MM approaches and 5) post-optimization strategies. Some examples are: the inclusion of Ir and Mn containing porphyrines into the Xylanasa A protein to construct and analyze the enantioselective profile of new peroxidases; [167] the engineered variant of Neocarcinostatin (NCS) developed by Mahy and coworkers for synthesizing asymmetric oxydases; [172] the study of the several ArMs synthesized by Roelfes and coworkers based on the LmrR protein bound to Cu(II)-phenanthroline and Cu(II)–bipyridine cofactors for the enantioselective hydration of ketones and the Friedel Craft alkylation reaction, or to hemin for enantioselective cyclopropanation; the Sav based ArMs developed by Ward's group which include a biotinylated Cp*Ir-Noyori's like catalyst for the asymmetric transfer hydrogenation of imines. [105, 166]

Integrative strategies has been demonstrated to be powerful not only for rationalizing experimental chemical outcomes but also to go one step further towards more intuitive frameworks related with the amazing area of enzyme design. Next Chapter summarizes the workline followed along the Ph.D. as well as the main goals which have driven the different studies comprised in this Thesis.

Chapter 2

Objectives

The main objective of this Ph.D. is the development of an efficient and robust computational framework for the rational design of Artificial Metalloenzymes (ArMs). This effort constitutes a natural step along the research line of the InSiliChem group as well as a major leap in moving from descriptive to predictive works.

Since 2010, our laboratory pioneered the development, optimization and application of multiscale strategies in Bioinorganics with a major focus to the study of key elements of the mechanistic features of ArMs (i.e. substrate and cofactor specificity or mutagenetic profiles). This Ph.D. embraces all the theoretical and technical knowledge reached so far and pretends to apply them to the computer aided **design** of novel enzymes.

Aware that this challenge can not be limited on pure computational inputs with no real world feedbacks, the work has been supported by strong collaborations with the experimental groups of Gerard Roelfes (University of Groningen) and Thomas Ward (University of Basel), two pioneers in the field of ArMs. In the following paragraphs, I present the specific objectives of the work and how they are divided between the different chapters.

Studies presented in Chapter 4, are aimed at testing several optimizations of our multiscale strategies. In particular, we expanded the workflow further than QM, dockings and QM/MM by introducing massive use of Molecular Dynamics. Focusing on two reactions, the hydration of alkenes by copper coordinating cofactors (Roelfes' group) and the reduction of cyclic imines by an iridium-piano-stool catalyst (Ward's group), our multiscale strategies have been substantially updated. This part of the Ph.D. present excellent agreement with experimental results including the understanding of directed evolution and regio-specific interactions. Thanks to these results, our updated framework serves as the starting material for the following works of the Ph.D.

Studies presented in Chapter 5, are focused on applying our skills to rational enzyme design and to test its viability. Focusing on the hydration reaction, we aimed at generating new systems using two different types of scaffolds: 1) the LmrR enzyme and 2) peptides up to 30 residues. The excellent results obtained overpassed the initial expectations. However, the set of studies also lead us to a series of unanswered but fundamental questions when dealing with the binding of organometallic compounds to protein, and the final chapter of the thesis concentrates on several of these aspects.

Finally, Chapter 6 aims at elucidating how a modeller can predict and handle possible complex conformational rearrangements after binding of homogeneous catalysts to proteins. As the mechanism of the binding of natural metallic cofactors like porphyrins has never completely been resolved and a long list of relevant heme proteins are available in the Protein Data Bank, we decided to start on those systems focusing on a variety of naturally occurring heme binding proteins. Surprisingly, in the process of this study, the group of Roelfes identified the strong binding of the heme group into the LmrR scaffold, which leaded to good cyclopropanation reaction profiles. Unfortunately, the solved X-ray structure is not able to characterize pre-catalytic binding modes of the heme and the substrates hampering possible re-design. Computation

has been employed to decode possible structural rearrangements of the ArM which allow the incorporation of the substrates for the reaction to proceed.

Chapter 3

Theoretical background

3.1 Molecular modelling

Theoretical chemistry aims at combining mathematical methods and fundamental laws of physics to study chemical processes. [173–176] The area which involves the resolution of the equations proposed by theoretical chemistry via computer simulation is known as *computational chemistry*. Previous to the development of electronic computers (1950's), the approachable calculations were quite limited. Over the decades, the improvements on the performance of computers have allowed the exponential increase of the complexity of calculations and subsequently the size of the systems that can be afforded. Nowadays, computational chemistry aims at assisting chemists by elucidating the transformations and properties of molecules for the description of chemical phenomena. This, in conjunction with the advances on the experimental techniques, have expanded the knowledge in different chemical subfields like transition metal catalysis and opened the door to areas of more recent enhancement such as enzyme modelling.

Although computers are the tool which allows applying theory to describe systems by performing complex calculations for us, they are just tools that still require hu-

mans to provide the initial description of the system by solving two main questions: 1) how the particles constituting the system are distributed and 2) what the nature of these particles is. The former regards the starting point accessible to the part of the phase space (all the possible values of the velocities and positions for all the particles composing a system, whose theoretical description is obviously out of context) which corresponds with a certain chemical phenomenon. The later relates to what is understood as "particles" since they can be considered of a very different nature: (bio)chemical processes are usually composed by a sum of tinny events occurring at different length and time scales, similar to a puzzle that have pieces which, at the same time, own specific shapes and draws. A colour blind person should be able to make the puzzle but fail to distinguish the range of colours which characterize each piece. Similarly, determined chemical events are only approachable by a determined theoretical method which will be dependent on the nature of the particles or "building blocks" (electrons, atoms, amino acids ...) selected to describe such phenomenon. For example, the conformational analysis of a molecule requires a method able to deal with the exploration of the conformational space in a reasonable period of time. In this case, atoms constitute the elemental particles of the system and the use of a level of theory which explicitly considers electrons would increase unnecessarily the calculation effort. In contrast, catalytic events involving bonds breakage and formation requires the analysis of a very specific conformation and a level of theory which explicitly accounts for the electrons, which in this case constitutes the elemental particles of the system. Concluding, depending on the type of chemical phenomena we want to describe, the different levels of theory must be consciously selected or combined to approach results at a reasonable accuracy/cost ratio.

Two main levels of theory are accounted for in molecular modelling: *Quantum Mechanics* (QM) and *Molecular Mechanics* (MM). QM explicitly considers nuclei and electrons of the molecules, so it is time consuming (being still limited to sys-

tems with no more than 500 atoms) but is able to model processes which encompass bonds breakage and formation. On the counterpart, MM considers atoms as classical particles with a specific mass, net charge and radius. Since it does not account for electrons explicitly, calculations are greatly simplified and thus, they can be applied to big systems and afford massive conformational sampling.

The next sections comprise an overview about the different levels of theory which constitutes the area of molecular modelling, as well as an outline of the available computational tools with particular focus in those used for modelling and describing the specific systems boarded along the different chapters of this Thesis. A very recent book which comprises the state-of-the-art molecular modelling has been recently published. [177]

3.2 Quantum Mechanics

3.2.1 Basic concepts

For an accurate description of the electron distribution QM solves the time-independent non-relativistic Schrodinger equation:

$$H\Psi(x) = E\Psi(x) \tag{3.1}$$

being the wave function Ψ related with the probability of finding these particles (electrons) for a determined quantum state. Ψ and E constitute the eigenfunction and the eigenvalue, respectively, of the Hamiltonian H, which constitutes the total energy of the system. However, only systems composed by one particle can be solved analytically. When applying QM to many-body systems an exact solution of equation 3.1 is not possible and approximations must be made. The first invoked one is the *Born-Oppenheimer approximation*. It bases its statements on the strongly

different time scales existing between the nuclear and electron motions. Due to the electrons are lighter than nuclei by several thousand times, they can be considered to move adiabatically around nuclei neglecting the existence of a finite relaxation time. Although, of course, exist some non-adiabatic effects that do not allow electrons to move instantaneously with nuclei, this represents an excellent approximation. Based on these assumptions, the Born-Oppenheimer approximation postulates:

- The wavefunction can be separated by the product of an electronic wavefunction and a nuclear wavefunction

$$\Psi_{tot}(R,r) = \Psi_{el}(R,r)\Psi_{nuc}(R) \tag{3.2}$$

- The electronic wavefunction depends parametrically on the coordinates of the nuclei.

Let us apply previous assumptions starting by writing the Schrödinger equation for a given molecule:

$$H_{tot}\Psi_{tot}(R,r) = E_{tot}\Psi_{tot}(R,r) \tag{3.3}$$

were the total Hamiltonian H_{tot} can be described as:

$$H_{tot} = T_{tot} + V_{tot} = (T_{el} + T_{nuc}) + (V_{ne} + V_{ee} + V_{nn}) \tag{3.4}$$

In this equation T_{el} and T_{nuc} are the electronic and nuclear kinetic energy, respectively, V_{ne} is the nuclei-electrons Coulomb attraction and V_{ee} and V_{nn} stands for the electronic-electronic and electronic-nuclei Coulomb repulsion, respectively.

Now, regarding that electrons are considered to move around fixed nuclei, 1) the nuclei kinetic energy term T_{nuc} can be ceased of being considered into the equation

3.4 and 2) the repulsion between the nuclei becomes constant. Thus, a new *electronic Hamiltonian* can be defined as:

$$H_{el} = T_{el} + V \tag{3.5}$$

which applied to the *electronic wavefunction*, results in:

$$(T_{el} + V)\Psi_{el}(R, r) = U_n(R)\Psi_{el}(R, r) \tag{3.6}$$

The resolution of equation 3.6 allows obtaining the electronic state of a molecule. U_n, also known as *adiabatic potential*, results from solving the electronic Schrödinger equation and constitutes the average value of the electronic Hamiltonian: the energy resulting from the sum of the electronic kinetic energy T_{el} and the potential energy of the nuclei V for a fixed nuclei coordinates. This dependence of U_n on the nuclear coordinates implies that for different nuclear geometries, we will get different U_n values and, consequently, different electronic wavefunction.

Last, by including U_n into the complete Hamiltonian, and applying the last to the nuclear wavefunction, it is possible to solve the complete Schrödinger equation:

$$(T_{nuc} + U_n(R))\Psi_{nuc}(R) = E_{tot}\Psi_{nuc}(R) \tag{3.7}$$

in which the complete Hamiltonian $(T_{nuc} + U_n(R))$ describes the motion of nuclei for an average electronic potential.

The calculation of the adiabatic potential U_n for a many nuclear coordinates constitutes the so-called *potential energy surface* (PES). The most interesting points into the PES and which are usually calculated in computational chemistry, are those related with stationary points, especially minima (minimum energy in all directions) and the transition states (minimum energy in all directions except one, which is a maximum) which relate these minima. The first provide us with useful information

such as what are the potentially observable intermediates during a reaction mechanism, and the later allows evaluating the reaction rate (Figure 3.1).

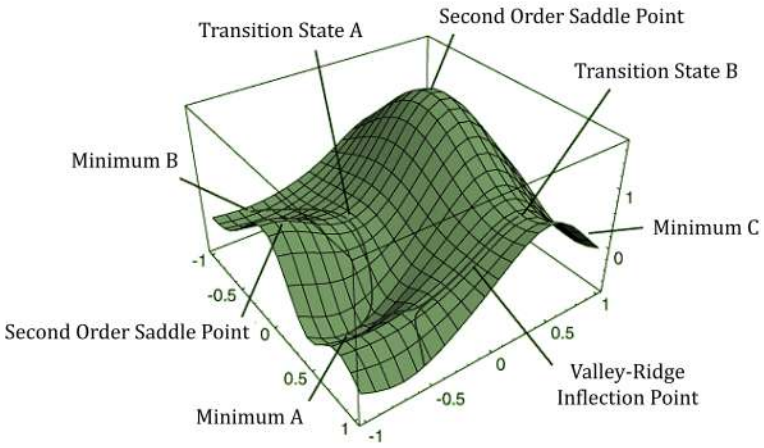


Figure 3.1: Graphic representation of the PES of an arbitrary molecule.

An additional level of approximations allows simplifying the resolution of the electronic Schrödinger equation. They can be separated in two main groups: the wave function based methods and the electronic-density based methods. Between the first group, the most widely used is the *Hartree-Fock* (HF) method, in which the electron-electron interactions are approximated in an averaged fashion. The model accounts for an exact electron-electron interactions or "exchange" i.e. the energy coming from the interaction between parallel spins of the electrons which tends to repel each other. In contrast, it neglects electron-electron "correlation", which is the instantaneous interaction between an electron and all the surrounding electrons, i.e. the spatial position of an electron is correlated with the spacial position of the rest of electrons. The inability of HF methods to account for this electronic correlation makes it to be poorly efficient to provide accurate chemical descriptions. This correlation can be incorporated by the help of *post-HF* methods but this still limited to small systems. Among the second group lies the *Density Functional Theory* (DFT), [178] which hardly improve the accuracy of the HF method by accounting for this electronic

correlation term. In this thesis all the calculations have been done based on DFT methods and thus, a more detailed description will be made in the following section.

3.2.2 The Density Functional Theory

Hohenberg and Kohn set out the basis for DFT in 1964 postulating that the ground state electronic energy can be determined by the electron density of the system (δ) [179]. This is, since the ground state energy is a functional of the ground state electron density, the different components of the ground state energy can also be determined. Density functional methods provide with significantly better results than HF since they account for electron-electron correlation by modelling it as a function of density. Additionally, while in the wave function based methods the complexity of calculation exponentially increases with the electron number (it contains 4N variables, three spatial and one spin coordinate, for N electron), in electron-density based methods the number of variables is the same (the electron density is the square of the wave function, integrated over N-1 electron coordinates, and thus the spin depends only on the three spatial coordinates) whatever the number of electrons. This allows the quantum study of relatively large systems with a good computational cost/accuracy ratio.

However, the fact that initial DFT methods express all the different components of the energy as a functional of the electron density has a consequence: the electron kinetic energy appears misrepresented, which resulted in an initial poor applicability of the method.

The Kohn-Sham theory

The success of the DFT in computational chemistry did not arise until 1965 with the establishment of the *Kohn and Sham* (KS) method, [180] which aims at finding a good approximation to the internal electronic energy of the system on its ground

state. The Khon-Sham method continues nowadays being the most popular for the electronic structure calculation in quantum chemistry.

It introduces a fictitious system composed of particles with same properties as the electrons composing the original many-body system, with the exception that it lacks the electron-electron interaction term, i.e. it describes a non-interacting system. These new particles move on a fictitious potential known as the *Kohn and Sham potential* defined equally to the system's ground state density. In other words, this assumes that the true ground state density is the same as the ground state density of a non-interacting system and thus, the calculation of the ground state wavefunction is far less complicated. The full wavefunction Ψ_S[1] can be simply calculated by placing individual single-particle wavefunctions, known as *Khon-Sham orbitals*, in a Slater determinant. The success of the method relies in the consideration of the individual orbitals. However, this has a double counterpart: on one hand, the increase of the complexity of the calculations from 3 to 3N variables, and on the other hand, that the electron-electron correlation becomes again a separate term.

The electron kinetic energy is obtained from the set of Khon-Sham orbitals placed in a Slater determinant used for representing the electron density in a non-interacting system as:

$$T_S = -\frac{1}{2}\sum_{i}^{N}\left\langle \varphi_i \left| \nabla^2 \right| \varphi_i \right\rangle \tag{3.8}$$

Since the kinetic energy T_S is calculated for a non-interacting system, it would be equivalent to that of the HF kinetic energy (just as HF orbitals describe non-interacting electrons).

Obviously this kinetic energy is not the same as that of the true system since, actually, electrons are interacting. So the necessary corrections are introduced into an *exchange-correlation term* E_{XC}, which is simply the fraction of the energy which

[1]The subscript S indicates that the term has been calculated from a Slater determinant.

remains after subtracting the non-interacting energy. The term $E_{XC}[\rho]$ corresponds to the *exchange correlation energy* and can be defined as:

$$E_{XC}[\rho] \equiv (T[\rho] - T_S[\rho]) + (V_{ee}[\rho] - J[\rho]) \tag{3.9}$$

where first parenthesis corrects the difference between the electron kinetic energy of the true system and the non-interacting system, and the second parenthesis corrects the difference between the defined electron-electron interaction potential and the classical one.

If E_{XC} was known, the KS method would give the exact energy of the system! Unfortunately, the weak point of the theory is precisely the resolution of this term i.e. the fraction of the kinetic energy corresponding to the electronic exchange-correlation. This represents the only unknown functional to achieve the desired coupling between the electron density and the exact energy of the system and, in its form relies the difference among the large variety of DFT methods.

Exchange-correlation functionals

Despite the large variety of functionals proposed for the resolution of the electron density function (several hundreds), a relatively small number is commonly used. As a consequence of their vast number, it is usual that the "oldest" and well-tested ones overshadow others that are more recent, sometimes due to their popularity or since the modern ones only provide slight improvements for specific cases and they are not usually included into the available programs.

When developing functionals it is common to separate the E_{XC} term into its components E_X and E_C, due to the different scalar properties that characterize these terms: the linear scaling of the electron coordinates result in the same linear scaling of the exchange energy (E_X), while the coordinate scaling of correlation energy

(E_C) is only linear at low-electron density. This implies difficulties to get correlation component that achieve cancelation of the long-range exchange term.

In an attemp to classify the vast variety of approximations of the $E_{XC}[\rho]$ functional, Perdew and Schmidt established different levels of approximations based on the well-known "Jacob's ladder of DFT" (Figure 3.2). [181]

Generally, these functionals can be grouped as *empirical* or *semi-empirical* if developed to fit experimental data for a set of particular chemicals or materials. In contrast, the *non-empirical* functionals are developed according to first-principle calculations. Empirical or semi-empirical provide with better results only when compared to non-empirical functionals of similar cost for the specific context they were designed for, but are very inaccurate when predicting properties for systems that were not present into the "training set". [182] In these cases, functionals based on first-principles appears more versatile to describe complex systems for which no experimental data other than fundamental constants is available.

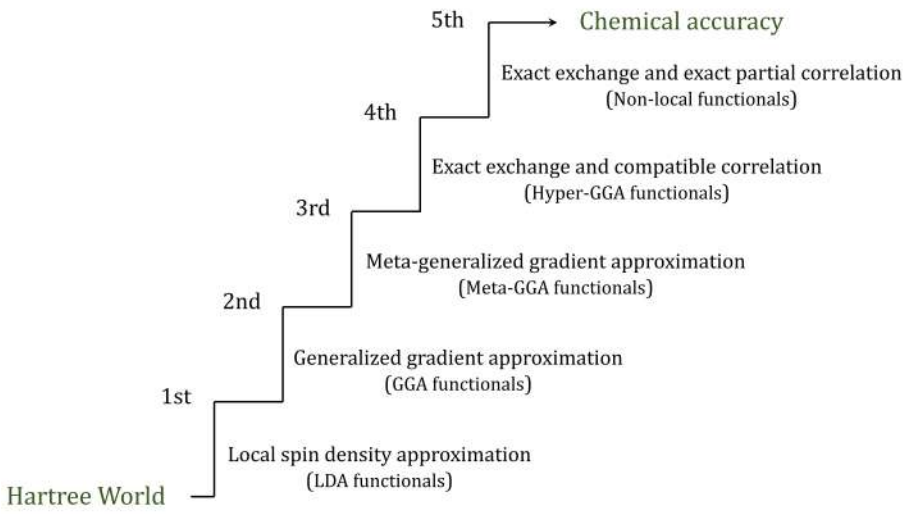


Figure 3.2: Illustration of the Jacob's ladder of DFT aproximations.

Next a brief description of the different exchange-correlation functionals will be performed by successively climbing the rungs of the Perdew's ladder:

Local Density Approximation (LDA) It assumes that the electron density can be treated as a uniform electron gas, and makes exchange and correlation energy dependent on this electron density. Variations for cases in which alpha and beta densities are not equal (that are quite frequent) have been included into LSDA functional, which consider individual densities for each type of electrons. It can be said that LSDA is an "exact" DFT method if applied to a uniform electron gas. However, for a molecular system, it tends to underestimate the exchange energy (leading to artificial stabilization of high spin states) and to overestimate the electron correlation by a factor close to two, which results in a poor description of the energetic properties.

Generalized Gradient Approximation (GGA) The improvement of the LSDA approximation relies on not considering the system as a uniform electron gas, but performing the exchange and correlation energy dependent on both the electron density and also on its derivatives. *A.D. Becke* proposed the B or B88 functional, [183] one of the most popular GGA methods which performs a correction of LSDA exchange energy (its error is reduced by almost a factor of two) by fitting experimental data. This exchange functional is frequently combined with LYP correlation functional provided by *Lee, Yang* and *Parr*, [184, 185] which parameters have been fitted to the He atom.

Other related exchange-correlation functionals are the OPTX [186, 187] and PBE. [188–190] The latter is a non-empirical functional result of the refinement of PW86 [191] and PW91, [192–194] all proposed by *Perdew-Burke-Erzehof*.

Meta-GGA Methods These group includes higher-order derivative variables, as the Laplacian of the electron density (second-order term) or the orbital kinetic energy (Laplacian of the occupied orbitals), being the later the more numerically stable although they both carry the same information. This inclusion of second order density information, leads to level-3 meta-GGA functionals ("pure" functionals), which

include the BR [195], B95 [196] (one of the few functionals which lacks the self-interaction problem), τ-HCTH [197], VSXC [198] and TPSS [199], being the later considered an improvement over the PBE.

The greatest improvement for chemical applications in terms of accuracy was during the 1990's with the inclusion of the exact HF exchange energy. This led to the level-4 hyper-GGA functionals (hybrid methods), briefly explained below.

Hybrid or Hyper-GGA Methods One of the major benefits resulting from the addition of HF exchange, is the reduction of the self-interaction problem, i.e. the electron density artificially interacting with itself. While in HF the definition of the Coulomb energy and the exchange promotes a total cancellation of this self-interaction term, in DFT this cancelation is not perfect. The key to perform an accurate functional is to include a suitable fraction of HF exchange, able to fulfill the deficiencies of the functional in such a way the errors coming from both pieces average to close to zero. The proper ratio of HF/DFT able to guarantee accurate results is still one of the trickiest points in the daily used methods.

Especially, a very delicate situation is when comparing close- and open-shell states, since the later appears over stabilized by the self-interaction error. In these cases, a functional with a proper HF contribution results crucial to prevent over or underestimation of the results. Unfortunately, it is unknown where this point lies. B3LYP, [200] [201] one of the most widely used functionals in computational chemistry is included in this group. It combines B88, 20% HF exchange and the LYP correction; plus three parameters obtained empirically, which allows the functional to properly fit experimental data. On the counterpart, as all standard DFT methods, it is unable to describe the non-covalent interactions related to dispersion forces, i.e. weak and cumulative short-interactions whose relevance increases with the size of the system. This issue can be solved by including the *Grimme's dispersion correction*

method. [202] To date, B3LYP-D3 (B3LYP functional plus dispersion correction) still represents one of the functionals with best global performance and is the one used for all the quantum calculations carried out along this thesis.

Variations of some lower order of the ladder functionals has been made by including the HF exchange leading to other hybrid methods of good performance such as B98, [203] τ-HCTH-hybrid, [197] PBE0 [204] or TPSSh. [205, 206]

Non-local functionals An additional level into the Jacob's ladder would be constituted by the so-called *Double Hybrid Methods*, which includes the full information of KS orbitals (DFT orbitals and orbital energies are used). Their main drawbacks are that they tend to perform systematic errors for the description of some properties, which, in conjunction with the increase of the computational effort, limits their general use.

3.3 Molecular Mechanics

When dealing with large systems (from thousands to millions of atoms), something frequent when working with biological macromolecules, it is not possible to perform a description based on full QM approaches. MM treats atoms as classical particles which behave based on the Newton's second law equation. [207] Electrons are not explicitly considered and, consequently, computational calculations are hardly simplified allowing the simulation of systems composed up to millions of atoms. The particles composing the system behave based on 1) pre-defined physical features (mass, charge and radius) and 2) the interactions between them, which come determined by the *force-field*.

3.3.1 General features on MM force-fields

A force-field is the result of solving the potential energy equiation (Equation 3.10). It aims to predict the energy of a system as a functions of its conformation and tries to reproduce structural properties of molecules based on data obtained empirically or via quantum calculations. This means that there is not an absolute correct form of a force-field, but it is specific for a determined description for which it is developed. The components describing the force-field include: force constants, atom charges and the Van der Waals radii parameters.

A good-quality force-field should be transferable through a set of similar molecules (for instance, parameters should be the same for modelling all n-alkenes), more especially if it is desired to make predictions. For this purpose, force-fields take advantage of the *atom type* concept, which considers that structural properties do not usually depend on each atom itself, but exist sets of atoms of same nature which behave similarly when embedded into similar chemical contexts. Thus, for example, all carbons contained into an aromatic ring behave based on the same parameters.

Nevertheless, there is still a fraction of cases in which an accurate work requires the specific parameterization of a molecule or a group of molecules. This is the normal case when working with proteins including non-common organic or organometallic moieties (such as ligands, organometallic cofactors or non-natural amino acids), which are not included into the standard biological force-fields.

The force-field energy is described as a sum of terms which represent the total energy required to distort the molecule for the given definition. They can be encompassed in two main groups, the *bonded* and *non-bonded* terms:

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross} \tag{3.10}$$

The bonded terms account for all atoms that are directly connected. These include: E_{str}, which is the energy function for stretching a single bond; E_{bend}, or the energy function for bending an angle (so it accounts for atoms separated by two bonds); and E_{tors}, or the torsional energy to rotate a bond (it accounts for atoms separated by three bonds).

The non-bonded terms describe "through-space" interactions between atoms that are not directly connected. They include both Van de Waals (E_{vdw}) and electrostatic (E_{el}) interactions.

Strong interactions coming from a specific term can promote certain structural distortion on the others. E_{cross} term accounts for these interactions by coupling the three first terms. This can include corrections of the type stretch-stretch, bend-bend, stretch-bend, stretch-torsion, etc. Different force-fields can differ in the type of cross terms they use (if any). As they are not crucial to describe structural properties they are only considered when requiring to model specific properties such as vibrational frequencies.

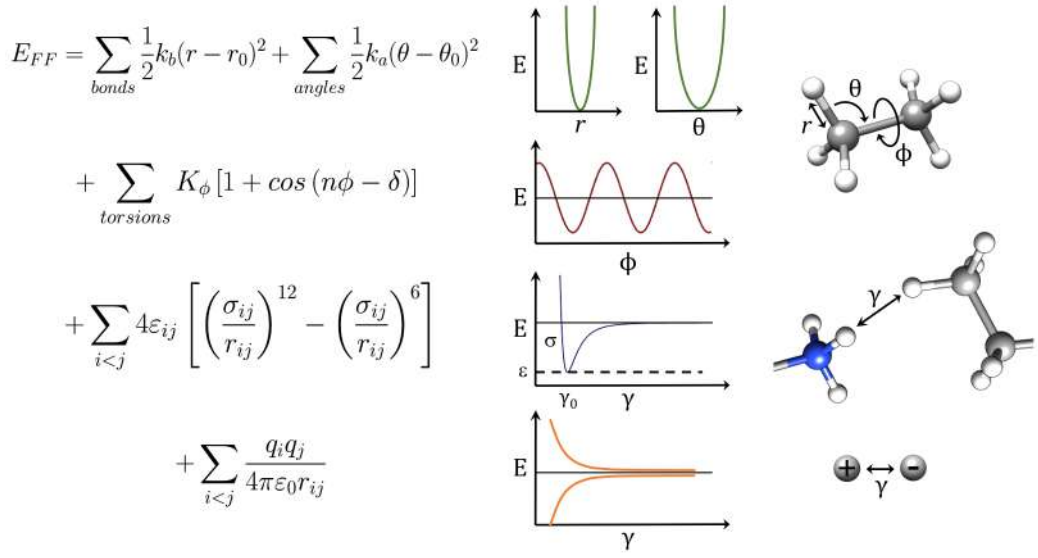


Figure 3.3: Illustration of the main descriptors of a general force-field.

Figure 3.3 offers a more detailed description of the above mentioned force-field terms. Energy related to *bonds* and *angles* follows a harmonic approximation in which the energy varies with the square of the displacement from the "reference" values[2] r_0 and θ_0 for bonds and angles, respectively; being r and θ the deviation with respect to this reference value.

Torsions are represented as a cosine series expansion which include the periodicity of the energy value; the multiplicity n represents this periodicity (being $n = 1$ for descripting a rotation periodic by 360°, $n = 2$ for periodicity of energy each 180°, and so on), the dihedral angle ϕ and the known as the phase factor δ, determines the point in which the torsion angle is a minimum.

In sp2-hybridized atoms exist a large energy penalty associated to piramidalization, since the four atoms tend to be in the same plane, where the equlibrum angle is zero. To account for the potential energy when the central atom is displazed out of the plane, usually an *out-of-plane bending* term (E_{oop}) is added, which is described as the quadratic function in the angle χ or the distance d (Figure 3.4):

$$E_{oop}(\chi) = k^B\chi^2 \quad \text{or} \quad E_{oop}(d) = k^B d^2$$

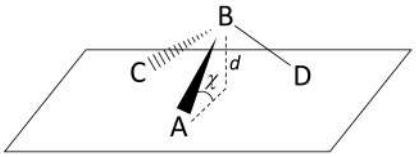


Figure 3.4: Out-of-plane bending variables.

[2]The "reference" value is the value of such bond or angle when the rest of the terms into the force-field are equal to 0. Notice that this differs from the "equilibrium" value, which is the value established when a minimum energy structure is adopted taking into account the contribution of the rest of the values conforming the force-field.

Regarding the non-bonded terms, the *Van der Waals* interactions are reproduced by a Lennard-Jones potential, being γ the distance between the interacting atoms, ε the Lennard-Jones well depth and σ the distance at which the non-bonded interaction becomes zero. With respect to the *electrostatic* interactions, they are described by a Coulomb potential between atoms i and j with partial charges q_i and q_j, respectively; ε_0 refers to the effective dielectric constant, which is 1 in vacuum and higher when existing solvent or intermediate atoms. As non-bonded interactions grow quadratically with the size of the system, they promote a rapid increase of the computational effort. To reduce this effect, a distance cut off is usually established by discriminating the long-range interactions beyond the assigned value. For further details the reader can refer to section 3.6.2.

3.3.2 Force-fields for biomolecular applications

Between the different force-field methods, the most commonly used for modelling large biomolecules are AMBER, [208] CHARMM, [209–219] and GROMOS. [220] All of them provide a fast computational performance by using quadratic Taylor expansions and neglecting cross terms, which allows dealing with big molecules in a reasonable period of time. Both AMBER and CHARMM are able to support proteins, lipids, carbohydrates, and nucleic acids. [221] In addition, they have been extended to describe a large series of common compounds in medicinal chemistry via the generalized CGenFF [222] (CHARMM) and GAFF [223] (Amber) force-fields.

In this thesis, the Amber force-field has been used to perform all the MM calculations. Several updates have been performed over time on the original Amber ff99, leading to new generation force-fields such as the ff99SB [224] and ff99SB*, [225] which include improvements on the backbone potential, and the ff99SB-ILDN force-field [226], which includes a modified side-chain torsion potential for a variety of amino acids. More recent developments includes the ff99SB-ILDN-NMR, [227] developed by

fitting parameters to experimental NMR data and the ff99SB-ILDN-Phi, [228] which improves the sampling in solvated models.

Although the different programs provide results with similar accuracy and have demonstrated to properly describe the molecular properties, their main limitation is that they do not account for changes on the electronic character along the simulation, i.e. they are not polarizable. Thus, they are limited to the conditions under which the parameters were developed. An alternative to this are the so called *polarizable force-fields*(Fluctuating Charge methods, ABEEM, Drude oscillators and Drude rod, induced dipoles, AMOEBA, PFF). [229] Although promising, accounting for polarizability involves considering additional degrees of freedom (which increases the computational effort), while still requiring additional parametrization. Thus, computational chemists usually still prefer to select the adequate standard force-field for describing the specific conditions under study.

3.4 Molecular Mechanics based approaches

Predicting the behaviour of biomolecules under certain conditions, as well as their interactions with their ligands and with other proteins, are between the most typical scenarios when working with enzymes. The modelling of such processes require a variety of approximations, including the use of MM force fields, which allow the treatment of systems of such dimensionalities. Among the most used approaches considering MM force fields are *Protein Ligand Docking* and *Molecular Dynamics* (MD) simulations. The former includes a much more simplified scheme than the later, regarding the nature of the force field and the algorithms implemented to define the molecular context. Going from the simplest to the most complex scheme, next sections will cover a brief description of these approaches, making emphasis on their use in the topic that concerns, the modelling of artificial enzymes.

3.4.1 Protein-ligand docking

Protein-Ligand Docking aims to provide the best arrangement between two interacting molecules, being the typical case a protein and a ligand (a substrate, an inhibitor, a catalyst ...). This approach has been extensively used in the area of drug design [230, 231] and, in particular, it is vital in virtual screening exercises in which large libraries of molecules are systematically docked into a protein binding site and evaluated in a short period of time. For that purpose, a good quality docking program requires two main features: first, to have a good compromise between a fast computational performance and enough accuracy for the evaluation of the docking poses; and second, to be able to reproduce experimental binding modes of the ligands. While accounting for those premises, all docking programs undergo two different stages: 1) sampling conformations into the protein binding site and 2) ranking the solutions based on scoring functions.

Sampling conformations at the binding site

Accounting for all the degrees of freedom of the protein and the ligand (six rotational and translational each), added to those related to conformational changes, the generation of all the possible binding modes ($\sim 10^5$) would be too demanding. To deal with a fast conformational search global optimization schemes based on sampling algorithms are usually employed.

Depending on the number of degrees of freedom considered, search algorithms can be categorized in three types: *rigid-body*, which only considers the six rotational and translational degrees of freedom; *flexible-ligand docking*, in which ligand experiences full or partial conformational flexibility at the protein rigid binding site; and *flexible-protein docking*, which, additionally, considers certain flexibility of the protein. Some subcategories arise depending on how to handle this ligand and/or protein behaviour. [232]

- A *systematic search* explores the majority of the ligand degrees of freedom.

- *Stochastic algorithms* explore the conformational space by changing randomly the ligand conformation. The Genetic Algorithm (GA) [233–235] is included into this group and is the one implemented in GOLD (Genetic Optimization for Ligand Docking), [236] the docking program used to perform all the docking simulations during this thesis. The name of the algorithm makes reference to the Darwin's theory of evolution, as it considers "genes" the degrees of freedom of the ligand and "chromosomes" the possible poses at the binding site. Then, the genetic operators *mutation* (random changes on genes) and *crossover* (change genes between chromosomes) act generating different poses that are iteratively assessed by the scoring function. After successive runs only best poses will survive throughout generations.

- *Deterministic methods* considers both protein and ligand flexibility by including Molecular Dynamic simulations and energy minimizations into the protocol. The problem of this strategy is that it is dependent on the starting point structure and is not able to overcome energy barriers to access to other local minima. [237]

One of the weak points of the standard docking approaches relies on their inability to consider an adequate adaptation of the protein binding site during the generation of the poses, since this requires dealing with an unaffordable amount of possibilities during the calculation. Some strategies aim to mitigate this problem by providing certain flexibility to the protein binding site:

- *Soft-docking* [238,239] aims to decrease the VDW radius allowing certain overlap between atoms. This is usually used in virtual screening approaches to expand the range of molecules able to fit at the binding site of a X-ray structure from which the inhibitor has been eliminated by hand.

- *Side-chain flexibility* [240, 241] allows the side chains of the residues located at the binding site to explore the conformational space with a limited number of torsional degrees of freedom which are based on pre-defined libraries of rotamers that predict low-energy conformations.

- *Molecular minimizations and simulations* of the protein-ligand complexes by different approaches such as Monte Carlo (MC) [242–244] or Molecular Dynamics (ensemble docking). [245] This has the advantage of generating accurate models accounting for all the conformational space and water molecules if necessary.

- *Collective degrees of freedom (CDF)* [246] approach allows accounting for different conformations of the receptor calculated via normal mode. [247–249] or by dimensional reduction methods such as principal component analysis (PCA). [250] This avoids dealing with high conformational variation while different conformations are taken into account.

An additional strategy implemented in GOLD for optimizing the ligand search procedure, is the identification of hydrophobic *fitting points* at the host binding site. For that purpose, a grid spaced by 0.25 Å is placed over the surface of the cavity and the Van der Waals interaction energy between each position of the grid and a bare carbon atom is evaluated. If this interaction energy is less than -2.5 kcal/mol, such point is automatically added to a list of fitting points. Then, during the conformational search, ligand lipophilic atoms are recognized and matched to this hydrophobic regions. If GOLD is not able to perform an appropriate map of fitting points for a specific system, it is possible to provide a customized MOL2 file which contains dummy atoms at the positions of interest to indicate GOLD where to locate them. The CCDC program SuperStar [251] helps to perform this customized file in the adequate format.

Evaluation of the poses

Although the advances on computation technology have boosted the quality of the search strategies, the evaluation of the generated poses still represents one of the weak spots of the docking technique. The reason is that there are still unsolved theoretical questions related to the enthalpic and entropic contribution to the binding affinity (Gibbs free energy), required to guarantee the proper discrimination among the correct vs. incorrect poses. [232] This evaluation is carried out by the so called *scoring* or *fitness function*, which quality is determined based on three main criteria: 1) the geometric docking accuracy, which measures the precision with which a native conformation of protein-ligand complexes is reproduced; 2) the screening utility, which aims at ranking the poses, properly identifying the potential ligands upon non-ligands between a pool of docked structures; and 3) the scoring accuracy, which aims to perform adequate rankings between a set of true ligands.

There are three main categories comprising the available scoring functions: [232, 252, 253] Knowledge-based, force-field-based and empirical approaches. Although all scoring functions comprise the same basic terms, they differ in the details of the functional form, the data sets in which parameters are based and the methods for optimimizing these paramaters.

- *Knowledge-based*: these strategies are based on the Potentials of Mean Force (PMF) [254] approach and use atom-atom distance potentials to fit contact preferences, which are derived from statistical physics. Their limitation is the amount of data available for the correct computation of the density distributions. The Astex Statistical Potential (ASP) [255] fitness function is an example.

- *Force-field based approaches:* the parameters for the different energy components are based on theoretical MM force-fields, and account for atom pairs interactions between the protein and the ligand described by a Lennard Jones

potential and a coulombic term. Their main drawback is that they do not consider properly the entropic effects. GOLDScore [236], the original scoring function in GOLD program, is an example.

- *Empirical:* in this case parameters are generated based on empirical data of ligand-protein complexes with known binding affinity values. The ChemScore [256] scoring function (Equation 3.11) is an example and it is the one used for all the dockings performed in this thesis.

$$ChemScore = \Delta G_{binding} + P_{clash} + C_{internal}P_{internal} + (C_{covalent}P_{covalent} + P_{constraint}) \tag{3.11}$$

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{hbond} + \Delta G_{metal} + \Delta G_{lipo} + \Delta G_{rot} \tag{3.12}$$

The ChemScore scoring function (Equation 3.11) value is obtained by adding to the total binding energy ($\Delta G_{binding}$) clashes penalty (P_{clash}) and internal torsion terms ($C_{internal}P_{internal}$). In case of covalent or restrained procedures, also covalent ($C_{covalent}P_{covalent}$) and constrained ($P_{constraint}$) terms are included. The total binding energy equation (Equation 3.12)) includes terms for the description of hydrogen bonding interactions (second term), ligand and host-metal ions interactions (third term), lipophilic host-atom/ligand-atom pairs interactions (fourth term) and the loss of conformational entropy after binding (fifth term).

3.4.2 Molecular Dynamics simulations

A MD simulation aims to simulate the behaviour over time of a molecular system at a finite temperature by solving the Newton's dynamical equation:

$$F = ma \tag{3.13}$$

Since nuclei are large enough, they can be treated as classical particles and thus, their evolution over time can be calculated by solving the differential form of Equation 3.14:

$$-\frac{dV}{dr} = m\frac{d^2r}{dt^2} \tag{3.14}$$

where V is the potential energy for a given set of Cartesian coordinates r. The left side of the equation corresponds to the negative of the energy gradient and relates to the force F acting on all the particles of the system. It should be calculated each time step and is the major responsible of the computational effort during the simulation.

For a given set of coordinates with current positions r_i, the positions of the particles for a later (Δt) or earlier $(-\Delta t)$ time step can be predicted by a Taylor expansion:

$$r_{i+1} = r_i + v_i(\Delta t) + \frac{1}{2}a_i(\Delta t)^2 + \frac{1}{6}b_i(\Delta t)^3 + ... \tag{3.15}$$

$$r_{i+1} = r_i - v_i(\Delta t) + \frac{1}{2}a_i(\Delta t)^2 - \frac{1}{6}b_i(\Delta t)^3 + ... \tag{3.16}$$

in which velocities v_i and acelerations a_i and hyperacelerations b_i are the first $(\delta r/\delta t)$, second $(\delta^2 r/\delta^2 t)$ and third $(\delta^3 r/\delta^3 t)$ derivatives, respectively, of the positions with respect to time t_i.

This algorithm, known as *Euler's* method, is a second-order expansion. This implies that each time step an error is accumulated leading to the continuous increase of the energy during the trajectory. Additionally, it lacks an important feature of Hamiltonian systems, which is reversibility, i.e. if you advance one step forward and then backward, you do not come back to the same point.

The *Verlet* algorithm [257] aims at modifying the previous equations to reduce the error accumulation and to make them time-reversible. Specifically, it allows solving the Newton's equation numerically and predicting positions a time step later or earlier from current and previous positions r_i and current acceleration a_i by:

$$r_{i+1} = (2r_i - r_{i-1}) + a_i(\Delta t)^2 + ... \tag{3.17}$$

$$a_i = \frac{F_i}{m_i} = -\frac{1}{m_i}\frac{dV}{dr_i} \tag{3.18}$$

Acceleration is calculated from forces F each time step (Equation 3.18), which then allow the propagation of the positions to generate the trajectory.

The main disadvantages of the Verlet algorithm are: on one hand, its proper formulation, which can lead to truncation errors since the new positions are the result of adding Δt (a very small number) to $2r_i - r_{i-1}$ (a very large number); and, on the other hand, that the velocity term does not appear explicitly. This is a problem since this term needs to be scaled when ensembles with constant temperature are generated. To overcome these issues, the *leap-frog* algorithm [188] modifies Equations 3.15 and 3.16 by considering half a time step as:

$$r_{i+1} = r_i + v_{i+\frac{1}{2}}\Delta t \tag{3.19}$$

$$v_{i+\frac{1}{2}} = v_{i-\frac{1}{2}} + a_i\Delta t \tag{3.20}$$

Although the numerical accuracy is better than in Verlet algorithm, the leap-frog algorithm updates the positions and velocities out of phase by half a time step. The *velocity Verlet* [258] algorithm aims to remove this abnormality by:

$$r_{i+1} = r_i + v_i \Delta t + \frac{1}{2} a_i \Delta t^2 \tag{3.21}$$

$$v_{i+1} = v_i + \frac{1}{2}(a_i + a_{i+1}) \Delta t \tag{3.22}$$

The calculated velocities allow determining the temperature more accurately and thus, getting a better energy conservation during the trajectory.

Thermodynamic ensembles

Microcanonical ensembles (NVE), for definition, aims to fix the total energy of an isolated system. However, in the real world, the system is subjected to energy exchange with the environment and, thus, they are not the best option when try to model chemistry or macromolecules. In contrast, in both *canonical ensembles* (NVT) or *isobaric-isothermal ensembles* (NPT) the system is able to exchange energy at a constant temperature with a heat or pressure bath, respectively, being the total energy equal to that the system plus the bath energy. The preservation of energy during dynamics is numerically difficult; however, this thermalization effect allows "instant" corrections of the particles for maintaining a constant temperature without altering the dynamics. For that purpose, heat bath adds or removes energy of the system by modifying the velocities of the particles, while pressure bath scales the positions of the particles promoting changes on the volume. *Nose-Hoover* methods [259, 260] have been demonstrated to properly deal with these both heat and pressure baths for producing true canonical and isobaric-isothermal ensembles.

The choice of using NPT or NVT to run our simulations depends basically on the purpose of our simulation. Specifically, when simulating the behaviour of biological macromolecules, the system should not be highly influenced by the choice of constant temperature or pressure if it is properly constructed. However, it is always preferred

to run the production at constant pressure (NPT), as this reproduce the experimental conditions.

Standard protocols usually initiate the simulation with, first, a minimization that help to relax possible steric contacts resulting from setting up the system and, then, an equilibration stage during which the system is progressively thermalized until reaching standard conditions (300 K). This stage is usually performed under NVT conditions to avoid artefactual expansion of the simulation cell coming from initial rearrangement of the system. Next, the production stage is performed at constant pressure (1 atm) and elongated in function of the molecular phenomena the user want to describe.

3.5 Hybrid methods

Modelling enzymatic reactions requires dealing with events occurring at different distance and time scales, making necessary the use of a variety of theoretical approaches. Hybrid methods appears as an elegant alternative to approach different simulation scales at once, allowing matching processes that, at first glance, contradict the theory explained so far, such as the possibility of performing a quantum study of the catalytic mechanism occurring into a such large system as a protein or the analysis of bond formation and breakage along a MD simulation.

Quantum Mechanics/Molecular Mechanics

As the use of quantum mechanical approaches is only affordable by limiting the length of the system, the precise study of reactions at macromolecular level appears out of reach. A solution to this is the multilevel strategy known as *Quantum Mechanics/Molecular Mechanics* (QM/MM) approach, which follows a "multilayered" strategy for the study of the reactions occurring into big molecular structures (usually

biological macromolecues). In other words, the system is managed at the inexpensive MM level of theory except in the region of interest, usually where a chemical reaction occurs, which is described at the expensive *ab initio* level. Thus, the partition of the total Hamiltonian can be expressed such as the sum of the MM and QM Hamiltonians plus a QM-MM interaction term as follows:

$$H_{total} = H_{QM} + H_{MM} + H_{QM/MM} \tag{3.23}$$

This constitutes a huge advantage since it allows the study of chemical reactions occurring in large (bio)molecular systems while maintaining a reasonable accuracy-time compromise. The combination of expensive QM approaches with a classical MM force-field was first stated by Honig and Karplus in 1971. [261] However, it was not until 1976 when Warshel and Levitt applied the QM/MM to an enzyme reaction and formalized the concept, [262] which led them together with Karplus and the entire series of results obtained during their fruitful career to be awarded with the 2013 Nobel Prize in Chemistry.

The hybrid *Our own N-layered Integrated molecular Orbital and molecular Mechanics* (ONIOM) method [263] developed by Morokuma and coworkers in 1999 is the one implemented in Gaussian program. [264] It allows the accurate study of big macromolecules by defining up to three layers or regions and has been tested to provide with comparable accuracy as if the whole system would be treated with expensive methods. [265] All the QM/MM calculations in this work have been done based on the ONIOM approach, so next paragraphs will be focused on how it deals with the QM/MM calculations.

Selecting the ONIOM layers

The ONIOM strategy aims to devide the system into different layers to enable approaching different levels of theory at once. Although for biological systems the most

common scenario is to divide the system in only two layers (QM:MM), it is able to assume up to three (QM:QM:MM) as described below (Figure 3.5):

- The chemical event occurring at the active region of the molecule usually includes bond breaking and forming events, so the only option is to treat it at high level. This region is set as the *high layer* and constitutes the Small Model system (SM).

- The entire system is included into the *low layer* of the ONIOM model, constituting the Real system (R), and aims at describing the environmental effects having impact on the events occurring at the high layer. It is described at a lower level of theory; usually MM, a semi-empirical method or an inexpensive *ab initio* approach.

- In some cases, a *middle layer*, which is normally treated with a middle accuracy between the two methods describing the others layers, is added for a better description of the effects occurring in the close surroundings of the SM. It is referred as the Intermediate Model system (IM).

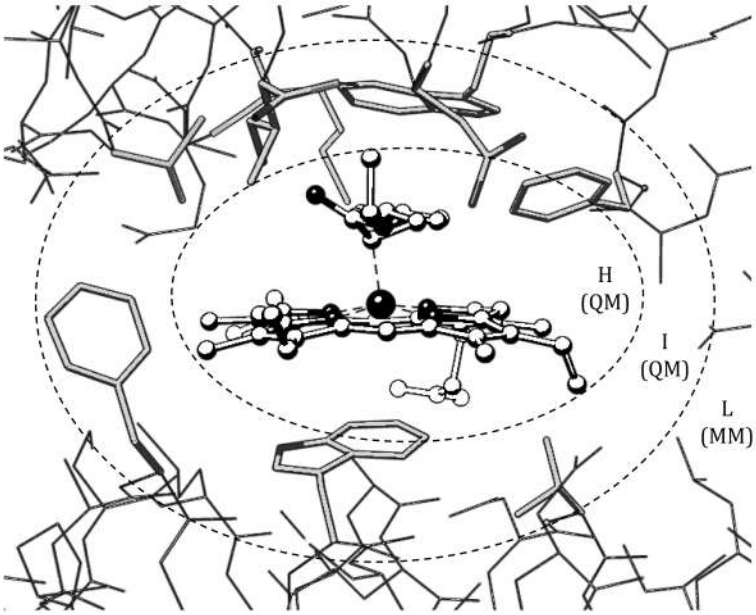


Figure 3.5: Illustration of the three ONIOM layers. The high system (H) is represented in balls and sticks, the intermediate system (I) in thick sticks and the low system (L) in narrow sticks.

Factors derived from the nature of the partition, such as bonds breakage or the nature of the interaction between the different layers could negatively affect the accuracy of the calculations. To prevent these issues, some technical considerations must be taken into account when setting up the ONIOM partition:

- The edge of the SM should be located at a minimum of three bonds far from any change on connectivity along the reaction, since MM parameters have to be the same for the reactants and products into the SM and R systems.

- Atoms connected by double, triple or highly polarized bonds have to be located into the same layer. Regarding the later, the presence of positive or negative charged MM atoms (such as N or O) may lead to attraction or repulsion of QM electrons promoting unreal distortions of the QM wave function. Thus, the partition should be done between non-polar bonds. When dealing with proteins, cutting between the alpha and beta carbons of the involved amino acids is usually a good choice.

The ONIOM extrapolation scheme

Supposing a two-layer model, as the one employed for the QM/MM calculations in this work, ONIOM employs a subtractive scheme (Equation 3.24):

$$E_{ONIOM} = E_{low}(real) - E_{low}(model) + E_{high}(model) \tag{3.24}$$

in which the SM is calculated at both low and high level, and the entire system R is calculated only at low level. Then, the SM low level result is subtracted from the total energy.

Describing the boundaries

The treatment of the boundaries between layers is one of the trickiest points during a QM/MM calculation as it deals with, on one hand, interrupted covalent bonds (if any) resulting from the stated partitions and, on the other hand, steric and electrostatics (polarization) effects between layers.

The cross of covalent bonds when defining layers implies the generation of a QM region with unpaired electrons, which have to be properly fixed. Different solutions are available to overcome this issue: One of them is the *localized molecular orbital* scheme, which aims at placing hybrid orbitals to saturate the QM region. [266] A more commonly used one is the *link atoms* scheme, which saturates the QM free valences with hydrogen atoms, only visible for the QM region, while the MM layer still considers parameters for the original bond. A variation of this may be applied when using semi-empirical methods by replacing this hydrogen by a pseudo-halogen atom with adjusted parameters to be as much similar as possible to the substituted atom. All of them are equivalent and provide with similar results if performed a correct layer definition, but the hydrogen link atom scheme is the most commonly used as it is the one which provide the easiest implementation. [267]

Regarding the QM/MM interaction term, it can be described through three increased in accuracy approaches: [268] 1) The *mechanical embedding*, which only considers the bonded and steric effects of the MM atoms on the high layer, while additional forces generated in both regions are included into the interaction term. Electronic interactions between the two layers are not taken into account and, although it is demonstrated that this does not necessarily have much influence on geometry, it can be important for the energetic description of the reaction mechanisms, especially if polar groups are present in the close surroundings of the high layer. An improvement to this is 2) the *electronic embedding* approach, which allows the electric field generated by the partial charges of the MM part to polarize the QM region by incorporating them into the QM Hamiltonian. This represents a major improvement for the treatment of the boundaries and only implies a small increase in the computational effort. Original ONIOM method only included mechanical embedding, but more recent versions have been incorporated the electronic embedding, which can be applied via the ONIOM=embedcharge keyword. Last, a more sophisticated approach is 3) the *polarizable embedding*, which also includes the polarization of the MM part by the QM atoms. However, this implies the use of a polarizable force-field and an iterative procedure which updates the electric fields of both QM and MM regions in a self-consistent manner. Apart from assigning partial charges to each atom (as electronic embedding), it also assigns dipoles, quadruple moments and dipole polarizability, leading to an unbalanced accuracy/time ratio which has displaced the use of the method.

What is the most adequate QM/MM framework?

Depending on the type of force-field and the basis sets selected QM/MM methods can be highly diverse. For enzymatic reactions it is common to use a fixed charge force-field for the R region and a DFT method with the desired basis set (normally

a medium-size one) for the SM system. It is stated that when using a fixed charge force-field the catalytic event occurring at the SM part should be placed at least at 10 Å from the MM region for a correct energy description. However, this implies an SM of around 1000 atoms, something unaffordable for an *ab initio* treatment. [269] In contrast, Fahmi Himo and co-workers have demonstrated, based on full QM cluster model approaches, that a good enough description can be achieved with QM regions consisting in only about 100 atoms. [270] It can be said that there is not a common agreement about a determined protocol to deal with QM/MM models, and that it is on the user hands to take the appropriate decisions to provide whit accurate results.

To further complicate the situation, a sampling problem exists when selecting the starting point structure (usually a X-ray structure or some frames from a MD trajectory). Selecting the adecuate starting point results critical since tiny changes on the MM layer can led to huge energy variations on the QM region. To reduce the margin of error as much as possible an intense pre-analysis and understanding of the catalytic event of interest results essential to approach its proper environment. Unfortunately, such amount of information is not always available and, in many cases, QM/MM results are not meaningful at quantitative level but they are only useful for qualitative descriptions.

3.5.1 *Ab initio* Molecular Dynamics

As in standard MD, in *ab initio* Molecular Dynamics (AIMD) atomic nuclei are treated classically and their motion can be computed by solving the Newton's second law. The main difference between the two methods is how the forces acting on the nuclei are calculated. As showed, in standard MD interatomic forces are obtained via a pre-defined potential energy, the force-field; in contrast, in AIMD these are obtained by calculating the potential energy surface each time step via QM. This is, the gradients of the energy resulted from solving the time-independent Schrödinger

equation with DFT (using the Born-Oppenheimer approximation) are then computed in function of the nuclear coordinates and used as interatomic potential to calculate the forces acting over the nuclei, which will behave again as classical particles.

This hybrid method results advantageous in the sense that, since potential energy is calculated via DFT, it allows the simulation of any non-parametrized system. However, it implies an additional computational effort which limits the size of the system that can be afforded.

The use of either *ab initio* or standard MD will depend uniquely on the features of the system the user want to describe. With AIMD, effects such as bond breakage and formation, polarization and charge transfer can be approached, while in MD this kind of effects come pre-imposed by the force-field and it is usually used for conformational exploration and statistical analysis. All the AIMD simulations performed in this thesis have been carried out with CP2K program. [271]

3.6 Solvation effects

Chemical reactions generally take place in presence of a specific solvent. For instance, biological reactions usually occur in water while, in (in)organic synthesis, the type of solvent employed (methanol, dichloromethane, dimethyl sulfoxide ...) can promote significant effects on the reaction outcomes, such as the reaction rates or the acid dissociation constants. This environmental influence on the molecular properties, requires accounting for an appropriate description of the solvation free energy into the simulations.

Based on their nature, the effects of solvation can be classified in: *non-specific*, which comprises the long-ranged effects promoted by the presence of the solute, such as solvent polarization and solvent electric multipole moments; and *specific*, which accounts for short-ranged solute-solvent interactions, such as hydrogen bonds, Van

der Waals interactions, the solvent shell structure, solvent-solute dynamics and charge transfer, hydrophobic and entropy effects. [176]

Depending on the strategy used to account for the energetic contribution of the solvent, different categories arise: 1) *Explicit* models consider discrete solvent molecules and all the interactions between them. They are computationally demanding but provide a spatially resolved physical description of the solvation interaction shell. However, under certain conditions, they fail at reproducing experimental data due to the use of fitting methods and specific parametrization when developing the models. A preferred strategy is the use of 2) *continuum* methods, which place the solute into a cavity and consider the rest of the environment by its average effect, i.e. as a continuum. They are more computationally economical while able to provide a reasonable energetic contribution and description of the solvent behaviour. Despite this, they fail at representing the local distribution of the solvent density around the solute cavity, an especially prevalent effect when using water as solvent. In addition, they are only useful if water molecules are not an active component along the catalytic pathway. An alternative to this is to use *discrete-continuum* strategies, which accounts for the first solvation shell by consciously locating a number of explicit water molecules at the close surroundings of the solute while considering the free space as a continuum. These models reduce the computational effort while achieving some spatial description of the solvent density around the solute. However, this treatment require additional knowledge to properly locate the water molecules around the solute and to decide how many they should be. Additionally, it is not recommended to add explicit water molecules when using implicit methods that have been developed by fitting experimental solvation energies.

Subgroups into the above categories emerge depending on the level of theory (QM or MM) at which the system is described. The majority of methods are focused on water since it is the most common solvent in biological media and also the most

difficult to be modelled. From now on, a brief summary of the strategies available to describe water solvent based on different levels of theory will be devoted, with an especial focus on water.

3.6.1 Explicit methods

This group provides with the most accurate representation of a solute embedded into a solvent as they account for specific solute-solvent interactions. Although under certain circumstances they can be useful in QM when requiring water as reactant or when performing cluster models, they are generally applied to MM, MD and Monte Carlo based approaches. While for the former the number of waters affordable appears highly limited by the size of the system, force field based techniques are able to account up to million waters molecules with a reasonable computational cost. Thus, modelling explicit water molecules is usually related with molecular descriptions and requires a proper sampling of the phase space via simulations. In this context, the developed water models aims to be simple enough to reduce the degrees of freedom to evaluate the solvent energy contribution while at the same time maintaining a good level of accuracy. On the counterpart, this simplification achieved by fitting parameters to experimental data makes some models to be useful only under certain conditions.

One of the most commonly used methods are the TPXP [272] (being X the sites of the molecule used to evaluate energy) and the simple point charge (SPC) models, which contain a fixed number of sites to analyze interactions (normally three for water) between molecules. Each site is described by a point charge, one dispersion and one repulsion parameters. Usually they are also bond and angle constrained. Between these, the three-point-charge based models TIP3P and SPC/E have been demonstrated to properly describe water properties for the liquid phase. [273] A more sophisticated strategy to describe the solute-solvent interactions, opposed to the above, is the use of polarizable force-fields, between which are AMOEBA (Atomic Multipole

Optimized Energetics for Biomolecular Applications), [274] SIBFA (Sum of Interactions Between Fragments Ab initio computed) [275] or QCTFF (Quantum Chemical Topology Force Field). [276] They are developed to reflect the charge anisotropy of the interacting molecules by accounting for multipole moments. Another promising polarizable water model is COS (Charge On Spring), [277] which allow the water molecules to be polarized by making flexible ("on spring") one of the interaction sites. Although being parametrized under standard conditions, it has been demonstrated to be efficient for describing solvent behavior at high temperatures and pressures. [278]

In addition to make a good choice regarding the water model, another concerning aspect when modelling (macro)molecules in solution is to avoid undesirable effects coming from the boundaries of the solvation cell. Generally, this is overcome by using the so called *periodic boundary conditions* (PBD).

3.6.2 Periodic boundary conditions

When modelling a solvated system, the incorrect treatment of the surface could promote undesirable effects during the simulation even considering an adequate number of surrounding water molecules. To minimize this issue, the molecule is located in the center of a water box, typically cubic (all types of space-filling polyhedral provide equivalent results), which is repeated in all directions constituting a quasi-periodic system with periodicity equal to the box dimensions (Figure 3.6). It is said then that the system is treated under periodic boundary conditions (PBC).

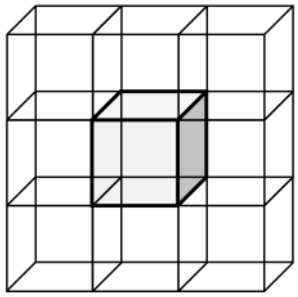


Figure 3.6: Periodic boundary conditions.

The major computational effort during a MD simulation comes from the calculation of the non-bonded interactions. Since their number increase exponentially with the size of the system, their calculation could be very time consuming when dealing with large molecules. However, this effort is not always necessary since some non-bonded interactions, such as Van der Waals, decay rapidly with the distance. Thus, usually a cut off distance is defined, which specifies an interaction sphere in which only atoms lying inside are considered. Since a direct truncation could promote negative effects (like a non-physical distribution of the water molecules) this is prevented via *switching* [279,280] and *shifting* [281] strategies, which aim at smoothing this effect by gradually powering down the Coulomb potential into a certain distance range. However, they both have the disadvantage of requiring pair-atoms lists each time step so they are still quite time consuming. Regarding electrostatic (Coulomb) interactions, they are long-ranged so they continue having relevance at large distances. As their effects would result ignored by the cutoff approximation, *Ewald sum* methods [282] are applied. They aim at splitting the generic integration potential into two summations: 1) the direct space sum, known as "near-field contribution", which represents the short ranged interactions and converge easily; and 2) the reciprocal space sum, known as "far-field" contribution. The famous *Particle Mesh Ewald* (PME) method [283] is basically a simpler way to calculate the Ewald summation through the application of Fourier transformations, which allow a significant decrease of the computational cost. This method also allows calculating the exact Lennad-Jones potential form of the Van der Waals energy [284].

3.6.3 Continuum methods

The preferred option for considering the solvent effects (when feasible) is the use of *continuum* methods. They are hardly simplified by embedding the solute S into a

cavity of a dielectric continuum with a specified dielectric constant ε, which defines the degree of polarizability of the solvent (Figure 3.7).

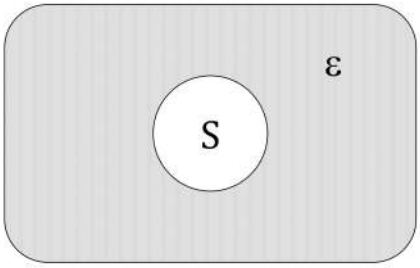


Figure 3.7: Continuum solvation model.

The charge distribution of the solute S polarizes the dielectric medium at the surface of the cavity. These charges act back modifying the solute's charge distribution, which defines a "reaction potential" (the polarization response of the solute to the solvent) that is iterated until getting self-consistency. Continuum models are frequently applied to QM strategies as these require reducing the degrees of freedom as much as possible. In this context, the charge distribution of the solute is iteratively calculated via *ab initio* calculations and is determined by the amount of perturbation the solvent performs on the QM Hamiltonian (Equation **??**) eq:QMSOLVENT

where Vmol+solvent is the term including the interaction operators required to describe the system response resulting from going from gas phase to a continuum medium. These interaction operators account for cavity generation, electrostatics energy, QM dispersion energy and QM exchange repulsion. [285] (rm) indicates that the equation is dependent on the solute coordinates. A variety of methods are comprised into the generic *Self-Consistent Reaction Field* (SCRF) term, which differ basically on how they deal with 1) the size and shape of the cavity, 2) the dispersion contributions, 3) the solute charge distribution and 5) the dielectric constant.

Between the routinely used continuum methods are those based in the Poisson-Boltzman equation, such as *polarizable continuum model* (PCM), and the *solvation models based on density* (SMD); [286] and those based on the Generalized Born (GB)

approximation,[3] like the *solvation models* (SMX). [285] The SMD approach has been used for all the QM calculations performed during this thesis in which continuum models have been applied. It considers the full electron density of the solute (without accounting for partial atomic charges) interacting with the medium. The cavity is defined as the union of interlocking spheres centered on each atom, accounting for parameterized radii and non-electrostatic terms. It has been parameterized based on solvation free energies for a large variety of solutes and solvents, being specially optimized for the B3LYP/6-31G* (the functional and basis set used for the QM calculations in this work) and the Minnesota family of functionals.

Another promising approach to model solvation effects is the COnductor-like Screening MOdel (COSMO). [287, 288] In contrast to the formers, which use an exact dielectric boundary condition, it uses a scaled-conductor approximation to derive the continuum polarization charges and approximates the cavity by segments to calculate the charges of its surface. The model have been demonstrated to be a powerful approximation and significantly reduces the error coming from the description of the outlying charge[4] with respect to the PCM method. [289]

The errors coming from lacking explicit interactions as hydrogen bonding appears to be more relevant when reproducing experimental data than the tinny differences coming from using different continuum models, which in general provide with a very valuable description of the solvation effects when proper cavities (close to the Van der Waals solute surface) are used. [289]

Regarding modelling macromolecules via simulation approaches such as MD or protein-ligand, this requires accounting for solvent effects as they are essential for describing solute reorganization resulting from changes on its polarity. This is quite frequent in daily studied events such as mutagenesis (protein design), ligand binding

[3] An approximation to the Poisson's equation applicable to any desired cavity shape

[4] This refers to the amount of solute electron density that reaches out of the approximated cavity, which can lead to significant errors in the solvent dielectric continuum treatment.

or protein-protein interactions. Even with a classical treatment, the all-atoms explicit approaches result too expensive for some applications, and implicit models gain again high relevance. In this context, the Generalized Born (GB) approximation is again the most widely used to capture the most important properties of the water solvent (polarity and polarizability). GB based approaches are able to properly represent continuum electrostatics, i.e. the solute atomic charges and polarized solving contributing to the solvation free energy, but fail at describing the non-polar contributions coming from solute-solvent dispersion interactions or exposed hydrophobic side chains. To overcome these issues, recent works focus on combining the polar contribution accounted by the GB approximation along with terms like the surface area (SA), Lazaridis-Karplus (LK) gaussian energy density, dispersion energy (DI) or Lennard-Jones-like non-polar contributions, which aim at including the non-polar component into the description of the solvation free energy and lead to combinations such as GBSA, GBDISA, GBDILK or GBLK. [290]

3.6.4 Discrete-continuum methods

The proper description of a system sometimes requires considering explicit interactions between the solute and the solvent, more especially in those quantum simulations in which water molecules are actively participating along the reaction mechanism. In QM and QM/MM based approaches it is frequent to resort to *discrete-continuum* strategies, which allow to account for discrete interactions of a small amount of properly located water molecules, while still maintaining a rapid performance by considering the rest as a continuum. A tricky point is to decide where to locate these water molecules (with which they are interacting and, in case of QM/MM, in which layer(s) they should be located) and how many they should be.

3.7 Integrative strategies

When setting up a computational framework for modelling processes occurring into an enzyme, it should account for aspects of very different nature, between which could be: 1) identifying adequate conformations to bind the substrate, 2) identifying a catalytically competent host-substrate fitting, 3) identifying the energetic barriers of the catalytic mechanism itself and 4) recognizing factors that could affect the chemical outcomes such as physicochemical changes into protein scaffold. Regarding a more specific case as could be the *de novo* design of an enzyme, the computational framework could additionally account for 5) the identification of adequate folds able to bind the substrate. If, moreover, the system consists in an ArM, also 6) the effects coming from the presence of the certain organometallic cofactor into the protein cavity should be considered.

Factors of such diverse nature cannot be described in conjunction by a particular molecular modelling technique, but they need to be rendered as separated molecular events. This leads to establish multi-scale strategies which aim at fractionating complex processes into particular events to be described with the adequate molecular modelling technique (Figure 3.8). In our group, integrative protocols combining QM, docking, MD and/or QM/MM techniques has been demonstrated to lead to a proper characterization of the conformational and catalytic properties of a variety of artificial systems. [95, 105, 166, 171] These have been particularly established for the study of ArMs and have provided with good experimental agreement.

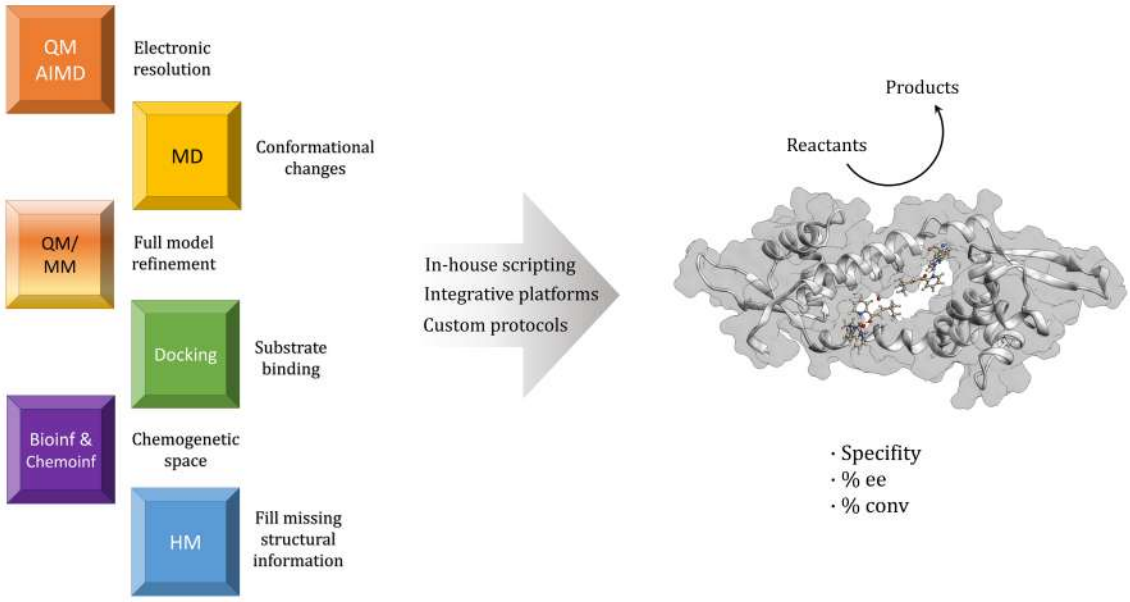


Figure 3.8: A computational chemist accounts for a variety of computational tools that must be properly combined into multiscale protocols to approach the description of the event of interest.

Combining several techniques to describe a whole molecular process is not an easy task and many difficulties arise from implementing integrative protocols. They come mainly from the necessity of generating an efficient flow between different programs which perform a different storing of the molecular data and whose inputs require specific formats. Some commercial interfaces exist which aim to integrate different molecular modelling techniques while avoiding these issues. However, they are not free and neither open-source. This means that there is not control on the algorithms driving the integrated steps, so performing changes for more specific purposes or integrating new functions cannot be boarded.

In our group many efforts have been dedicated to the design of free available integrative interfaces taking advantage of the open-source program UCSF Chimera, [291] developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311). It is easily extensible thanks to the existence of glue code written in Python that

allows coupling of non-commercial software and/or embedding new projects into the Python interface.

3.8 Dealing with metals in molecular modelling

Dealing with systems containing transition metals require in most cases considering electronic effects since they usually participate in processes involving changes on their oxidation or spin state, ligand coordination exchange (including reactants or/and binding amino acids) and/or variations in the coordination geometry. [229] This generates an extra difficulty during the modelling process, which has promoted a generalized refuse by the theoretical scientific community to study or design systems merging transition metal catalysis and biochemistry (like ArMs). Proof of this is the reduced number of groups existing worldwide which deals with the *in silico* modelling of ArMs.

3.8.1 Metals in Quantum Mechanics

The QM calculation of transition metal containing systems appears as a major problem due to the near degeneracy associated to the electrons that are partially occupying the *d* orbitals. To obtain very accurate calculations, multireference methods like the Multi-Reference Configuration Interaction (MRCI) are required. [229] However, these are computationally very expensive and, thus, computational chemists are more interested in identifying cheaper computational methods that are still able to provide a good performance. In this context, DFT methods appears promising as they are able to describe the electron correlation effects while owning advantageous scaling properties with respect to post-Hartree-Fock methods. This makes DFT to be widely used for the computational study of large compounds containing transition metals, including also the *ab initio* method used as part of the study of large systems such as DNA

or proteins. [292] Between the most popular functionals for describing the structural and energetic properties of transition metals containing systems, the hybrid-GGA functional B3LYP, [200, 201] together with BP86 [183, 191] and PBE0, [204] shows a generalized good performance in many benchmarks based on different sets of experimental data, including databases of general use for functionals development. [293–295] Especially for organometallic compounds, hybrid functionals containing more than 40% of HF exchange are not recommended. [229] Additionally, for GGA, meta-GGA and most hybrid-GGA functionals it has been found that the magnitude of heat of formation error highly varies in function of the degree of coordination of the metal, in some cases achieving error values higher than 55 (kcal/mol). [296] Specifically, B3LYP have found to be one of the few for which this error is maintained stable (less than 15.0 kcal/mol) whatever the degree of coordination is.

3.8.2 Metals in Molecular Mechanics

As mentioned in previous sections of this chapter, MM techniques such as docking or MD simulations perform the energy description of the system based on specific force-fields of different nature. The main limitation for describing organometallic complexes via force-field based approaches is that the model appears limited to discrete states of the metal-ligand interactions.

Regarding docking simulations, their force-fields have not been designed to consider interactions between the protein and the first coordination sphere of organometallic ligands, as they were originally developed to approach screening processes in which only organic pharmacological compounds were considered. However, the increasing interest in the last decades in the metalloenzyme design field has boosted the necessity of dealing with modelling the metal coordination geometries. A variety of strategies able to mitigate this limitation have been reported. Some are based on the automated calculation of the metals interaction geometry (implemented

in FlexX software), [297] and others based on the Cationic Dummy Atoms Model (CDAM), such as the "H-bond trick" strategy, [298] successfully implemented in GOLD5.2 program. [236] This places atoms acting as hydrogen bond donors at the coordination sites of the metal. Another recently developed strategy that have been included into the homology modelling program MODELLER [299, 300] consists on applying restrains to satisfy correct distances, angles and torsions between the metal and its environment via a geometric optimization scheme. [301]

An even more challenging scenario is faced when performing MD simulations of biomolecules containing metal moieties, since generalized force-fields applicable to biological systems does not account for parameters describing metal interactions. Due to the large variety of structural rearrangements among metals and their coordinating ligands, accounting for all of them would be absolutely out of range. Force-fields like AMBER or CHARM include extensions of parametrized organometallic compounds frequently found into modelled proteins, such as an iron bound to a porphyrin and a coordinating histidine. However, they would not be able to provide a proper description of the same catalyst included into a different protein context. Usually, when working with organometallic motifs is in the user hands to ensure their proper description by feeding the force-field with new parameters.

For an appropriate description of the metal interactions, getting bonded (distances, bonds, angles) and non-bonded (electrostatic, Van der Waals) force constants requires either experimental data or *ab initio* calculations which provide with proper vibrational frequencies between the different sets of atoms. Then these can be translated to representative force constants throughout different strategies, being the most popular the Seminario's method [302] due to its easy use and its good accuracy. It has been implemented into programs such as MCPB [303] and VFDFT [304] to facilitate its application. Regarding charges, they can be calculated via different approaches. The most popular is the *restrained electrostatic potential* (RESP) [305] method, which

generates atom-centered point charges based on an algorithmic charge fitting procedure to an electrostatic potential (ESP) generated via QM. Other empirical/semi-empirical and cheaper methods are AM1-BCC, [306] CM2, [307] Mulliken, [308] or Gasteiger. [309,310] All of them together with RESP are accessible via the antechamber program. [311]

An alternative to the above is the already mentioned CDAM approach, [312] which aims to mimic the interactions of the metal by consciously locating positively charged pseudo atoms at the metal coordination sites. In contrast to the above strategies, the CDAM method is independent to the context and thus, transferable. The model was firstly introduce for Mn^{2+} but was then extended to a large variety of different metal centers. [313]

Although some force-field able to consider electronic effects have been developed (YETI, [314] SIBFA, [315] LFMM [316]) and has been successfully implemented in a variety of studies involving organometallic compounds (some even into the context of metalloproteins), [317–319] these are far from being extended to the enzyme design field do to the difficulty on their application. Instead, computational chemists prefer make use of standard force-fields and expand the computational framework to describe the system under QM including integrative strategies, as is the case of the work performed in this thesis.

Chapter 4

Rationalizing ArMs via multi-scale computational frameworks

The computational study of ArMs is a complex task since it involves dealing with complex processes, ranging from recognition phenomena to catalysis, most of which are mediated by the metal. The Insilichem team has focused on studying these systems during the last 10 years through the implementation of multi-scale strategies.

In previous work performed in the group it was observed than one of the major limitations for ArM design is the large geometrical variability of the homogenous catalyst inside the receptor before and/or during catalysis. From a computational perspective, this makes the search of pre-catalytic structures (the ones that are in convenient molecular configurations for the reaction to proceed) particularly challenging because of the extensive exploration of the conformational space that is required.

Due to the difficulty of setting up a framework able to describe metal mediated pre-catalytic states at large scale, the previous studies performed in the group only considered quantum calculations (QM and QM/MM), structural statistics and fine tuned Protein-ligand docking simulations as part of the integrative strategy. [166] This framework appears adequate enough to characterize physically relevant starting

points but, depending on the system under study, it is quite limited to access to pre-catalytic conformations. In this sense, MD simulations appear of great interest in the way that they could potentially refine the initial Protein-ligand docking candidates through the exploration of the conformational space.

One of the ground objectives of this Ph.D. has been to embed classical MD simulations as part of our integrative strategy. For this purpose, it was fundamental to 1) implement and optimize metal parametrization for generating force fields of the metal complexes and 2) use at its maximum expression the accelerated drivers that have been appearing in the computational panorama in the last 10 years (especially for GPUs).

In the work presented in this Chapter, computational frameworks including MD simulations have been used to characterize the mechanisms occurring into a variety of ArMs. Specifically, these has been implemented to rationalize the entantioselective tendencies of first, an artificial hydrogenase designed by Roelfes group (The University of Groningen) and, second, three variants of an artificial cyclic imine reductase optimized by Ward group (The University of Basel) throughout directed evolution approaches.

These studies represent the cornerstone of the work developed in the following parts of the Ph.D. and more particularly in setting up the computational protocol for our main goal: the computational *de novo* design of artificial systems.

4.1 The study of the Cu-catalysed hydration of alkenes in water and into the context of an artificial metallohydratase

Chirally pure alcohols are transversal intermediates in chemical industries with a wide range of applications spreading to fine chemistry, perfumes and cosmetics. [320, 321] One of the most interesting reactions to approach those compounds is the enantioselective addition of a water to alkenes, a mechanism that provides atom economy and biocompatible processes. The base-catalysed 1,4-addition of water to unsaturated double bonds (Michael addition) [322] is a text book reaction known to occur more efficiently in electron-deficient C=C bonds. [323] Biocatalysis could represent an interesting option in this aim. Some naturally occurring enzymes like fumarase and enoyl-CoA hydratase have attracted the attention of chemists due to their excellent catalytic and enantioselective profiles. [324–327] However, they are extremely specific in term of substrate scope and their biochemical modification for industrial purposes is still very limited. [320, 328, 329] Only few examples of efficient naturally occurring enzymes able to perform the hydration reaction are described. [330, 331] Still, they proceed with very low ee levels. The design of an enantioselective hydratase with broader specificity to reach chemically relevant reactants appears therefore as an interesting goal. A promising strategy to address this challenge is the use of Artificial Metalloenzymes (ArMs) to drive the activation of the water nucleophile and induce asymmetry by the second coordination sphere. [87]

To date, the most successful ArM design for the asymmetric hydration of alkenes has been reported by Roelfes and coworkers. [94, 95, 332] These are based on a copper containing catalyst (phenanthroline or bipyrine) embedded in different biomolecular scaffolds, either DNA or protein, by supramolecular, covalent or artificial amino acid anchoring. Despite those advances, the control of Cu(II) mediated hydration still

presents a series of challenges to develop better biocatalysts. The main difficulty lies in dominating the reactivity of water molecules that are, at same time, a small asymmetric reagent as well as part of the biocompatible media in which the reaction takes place. Despite water has been clearly identified as a nucleophilic agent for the reaction, the mechanistic picture of enantioselective Cu(II) mediated Michael addition is still very partial and full of hypothesis. Scientific community points towards a stepwise process involving nucleophilic and electrophilic attack in an *anti* fashion, in which the final enantiomeric excess (ee) is defined in the nucleophilic attack.

In this Chapter, we will study the ArM designed by Roelfes and coworkers, based on the inclusion of a copper containing phenanthroline cofactor (Phen-Cu(II)) at position M89C/M89C' of the dimeric LmrR protein (Figure 4.1). This artificial matelloenzyme was able to perform the enantioselective addition of water to ketones with high enantioselective levels. According to experimental data, the best results were obtained for the substrate **2b**, which corresponds with an α,β-unsaturated ketone containing a t-butyl as R substituent. Interestingly, deuterion experiments on the isolated Phen-Cu(II) catalyst showed that the stereoselectivity comes from the catalyst-substrate complex itself, which favors a *syn* electrophilic attack (REF). With respect to the enantioselective character of the artificial metallohydratase, mutational analysis of the LmrR scaffold suggested that the critical amino acid for the efficiency of the enzyme was the residue D100, without which both conversion and enantioselectivity drastically decreases. [94] This is consistent with evidences of a variety of hydratases which are able to use conveniently placed negatively charged amino acids for the direct or indirect activation of the water molecule. [328, 329, 333, 334]

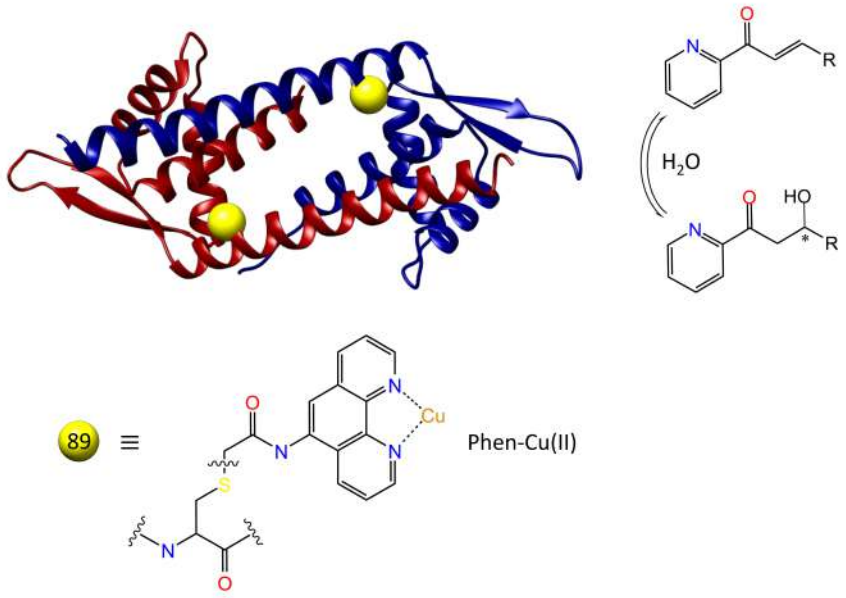


Figure 4.1: Illustration of the LmrR based artificial metalloenzyme designed by Roelfes and coworkers for the enantioselective addition of water to ketones. Two Phen-Cu(II) cofactors were covalently linked to positions M89C/M89C', aiming to pack them at the LmrR hydrophobic dimer interface.

This work aims at providing a clearer description of 1) the catalytic mechanism of the copper mediated hydration of ketones in water and 2) how the LmrR scaffold modulates the catalytic process to achieve enantioselectivity. For this purpose, we perform a multiscale computational framework which spreads over a large variety of methods (Figure 4.2). This combines Quantum Mechanics (QM) calculations, *ab initio* Molecular Dynamics (AIMD), Protein-ligand docking, Molecular Dynamics (MD) simualtion and Quantum Mechanics/Molecular Mechanics (QM/MM). Specifically, the process consisted in:

- Characterizing the copper mediated hydration of ketones via QM cluster models (Figure 4.2 a). The entire mechanism was assessed with a cluster composed of the Phen-Cu(II)-**2b** complex plus six water molecules. Critical points such as the distribution of water molecules around the catalyst-substrate complex and the discrimination between a step-wise or concerted mechanism, were further

assessed via *ab initio* MD simulations (AIMD). The inclusion of water molecules into the cluster model limits the study of the stereospecific nature of the reaction. The required displacement of water molecules to approach the substrate in *syn* or *anti* would lead to non comparable intermediates along the reaction pathway. Thus, an alternative cluster composed of the Phen-Cu(II)-substrate and a pyridinium (proton donor) was carried out to compare between the *syn* and *anti* energetic barriers. Last, we wanted to assess the effect of using an aspartate residue as water nucleophile activator. For this purpose, a third cluster composed by the Phen-Cu(II)-substrate complex plus six water molecules and an aspartate moiety was performed.

- Assessing the assembly between the LmrR and the Phen-Cu(II) cofactors. A system composed of the LmrR protein with Phen-Cu(II)-substrate and Phen-Cu(II)-2H_2O complexes included at positions 89 and 89', respectively, was constructed via Protein-ligand docking. This system was analysed through large scale MD simulations (see details in next section) ((Figure 4.2 b)).

- Studying the role of the second coordination sphere of LmrR as well as identifying pre-catalytic frames ((Figure 4.2 c)).

- Characterizing the copper mediated hydration reaction into the LmrR protein via QM/MM and full QM cluster calculations (Figure 4.2 d).

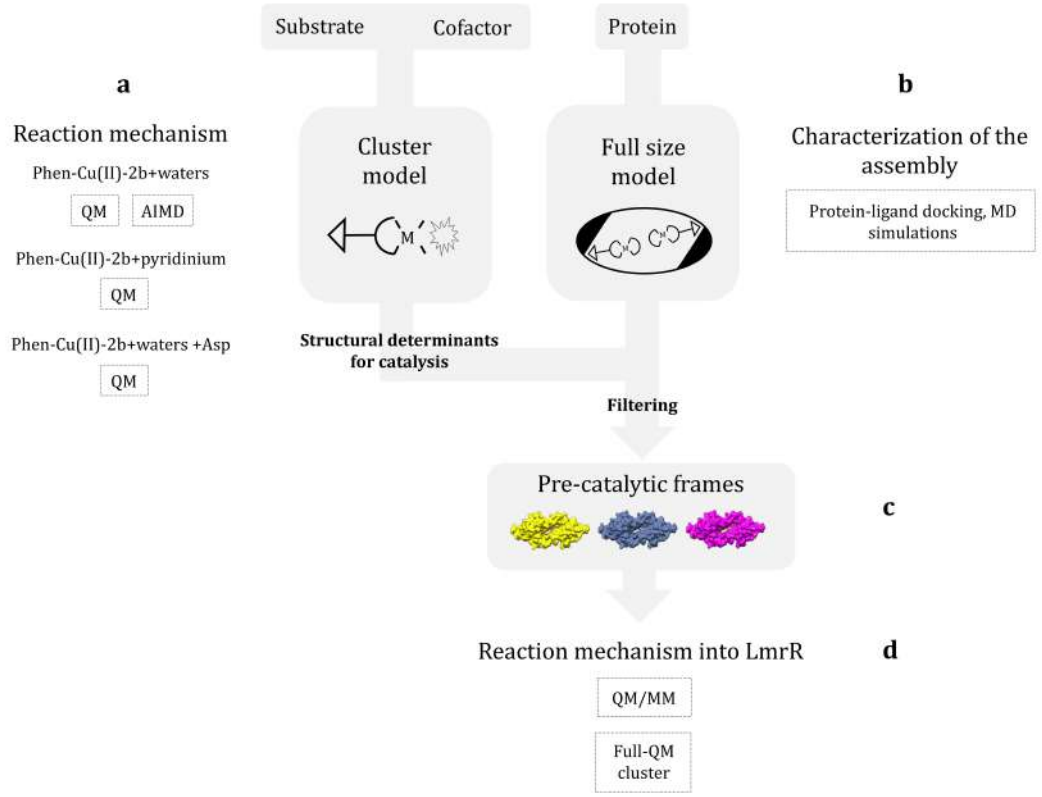


Figure 4.2: Computational framework for the study of the copper mediated hydration reaction implemented in this work. This include a) QM cluster calculations to characterize the hydration mechanism, b) the characterization of the full size model including the Phen-Cu(II) cofactors linked at position 89/89' via Protein-ligand docking and MD simulations, c) the identification of pre-catalytic frames involving the second coordination sphere of LmrR and d) the study of the hydration mechanism into the context of the LmrR scaffold.

This integrative approach sheds light on the copper driven hydration mechanism as well as the effect coming from the second coordination sphere of LmrR. We believe this work opens major avenues for both organometallic and biocatalysis fields, as well as elucidates skills for optimizing the efficiency, regarding both conversion and enantioselectivity, of artificial hydratases.

4.1.1 Computational methods

Quantum calculations

All the quantum calculations were carried out with Gaussian09 program [264] at DFT level using B3LYP-D3 functional. [200–202] The 6-31G(d,p) basis set was used [335–337] for non-metallic atoms and SDD [338] including a f polarization function for the 19 outer electrons of copper (the 10 inner core electrones are described by the SDD effective core potential (ECP)).

The implicit Solvation Model based on Density (SMD) was used as continuum method. [286] For those cluster models containing second coordination sphere amino acids, an epsilon adequate to describe solvent accessible surfaces of proteins (ε=9) was specified. [339]

Ab initio Molecular Dynamics

AIMD simulations were used to investigate the structure of the first solvation shell around the nucleophilic water and the substrate double bond. The entire process appears summarized in figure 4.3. In particular, we wanted to assess:

- The existence of proper water chains (composed of 2, 3 or 4 water molecules) connecting the nucleophilic water and the double bond of the substrate. This search was performed along 12 ps of restrained AIMD, in which the atoms of the nucleophilic water were restarined close to the substrate double bond (Figure 4.3 b). To identify water chains along the AIMD trajectory, we used an adapted version of the algorithm proposed by Stirling and coworkers, [340] which is able to identify up to three member water chains between two specified points.

- The destination of the proton after releasing the restrains taking as starting point periodically selected frames (each ps along the last 6 ps) from the restrained AIMD trajectory (Figure 4.3 c). Additionally, frames that contained

water chains, identified in the previous step, were also considered. All of them were selected and submitted to 1 ps of unrestrained AIMD to assess if the destination of the H+.

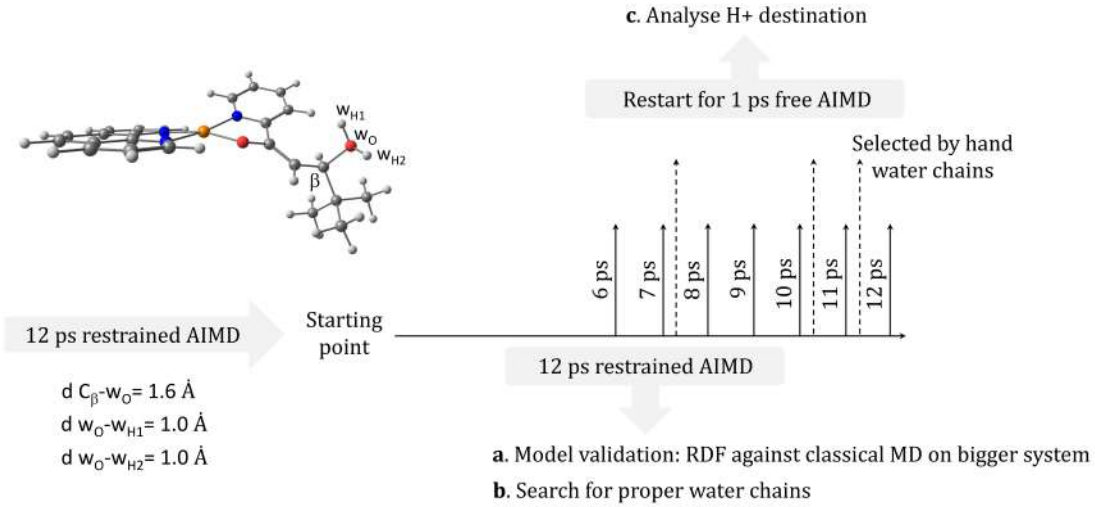


Figure 4.3: Workflow to study the nature of the proton jumping via classical and *ab initio* MD simulations.

The optimized structures were solvated with TIP3P 228 water molecules in a cubic box of size L = 19.302, so that reproducing the density of water at ambient conditions. Initially, this small size system was submitted to 12 ps of classical MD simulations under periodic boundary conditions using sander simulation program [341, 342] to equilibrate water molecules around the organometallic complexes. The distances between atoms into the nucleophilic water molecule (w_O-w_H = 1.0 Å) and the water molecule and the substrate (w_O-Cβ) were fixed in order to avoid the proton hopping at this point. The final conformation of the simulation was used as starting point for 12 ps of AIMD.

All the AIMD simulations were carried out with the CP2K program [343] under NVT conditions. The force constants were calculated each time step at DFT level using the BLYP functional [183–185] and the DZVP-MOLOPT-SR-GTH basis-set. [344] The core electrons were described based on pseudopotentials. [345] The equations

of motion were integrated with a time step of 0.5 fs using the Nose–Hoover thermostat. [259, 260] At this stage, the mentioned distance constraints were maintained.

Aiming to verify that the dimensions of this small and charged box was able to properly represent the distribution of the water molecules around the system, the AIMD simulation was compared to an equivalent classical MD of the same system embedded into a larger and neutralized simulation cell. Comparisons were based on the radial distribution functions of both the nucleophilic water and the carbon acceptor of the proton (Figure A.1). See details for classical MD below.

Protein-ligand docking

The crystal structure of LmrR bound to the drug daunomycin (PDB code: 3F8F) was used as initial model. Both daunomyicin and crystallographic water molecules were removed. To generate the ArM two different ligands were constructed: 1) Phen-Cu(II)-**2b** and 2) Phen-Cu(II)-2H_2O. Final geometries resulting from optimization via QM calculations were used for covalent docking at position 89/89'. A bi-coordinated geometry of the substrate at the equatorial position with respect to the copper cofactor was considered, since this appears to be the most adequate to fit the active site (Figure 4.4).

For the inclusion of both complexes into LmrR, both M89 and M89' were mutated to C using the Dunbrack rotamer library [346] implemented in Chimera program. [291] The covalent linkage was performed between the sulphur atom of the cystein and an overlapping sulphur atom added to the phenanthroline linker. Two successive docking runs were performed: first, the complex Phen-Cu(II)-**2b** was included at position M89C; second, Phen-Cu(II)-2H_2O complex was docked at position M89C' of the resulting structure from previous step. For a proper disposition of the substrate at the dimer interface, the complex 1 was selected as the first docked structure.

The dockings were performed with GOLD 5.2 program (available through the Cambridge Crystallographic Data Center (CCDC)) [236] and results were evaluated using the ChemScore [256] scoring function.

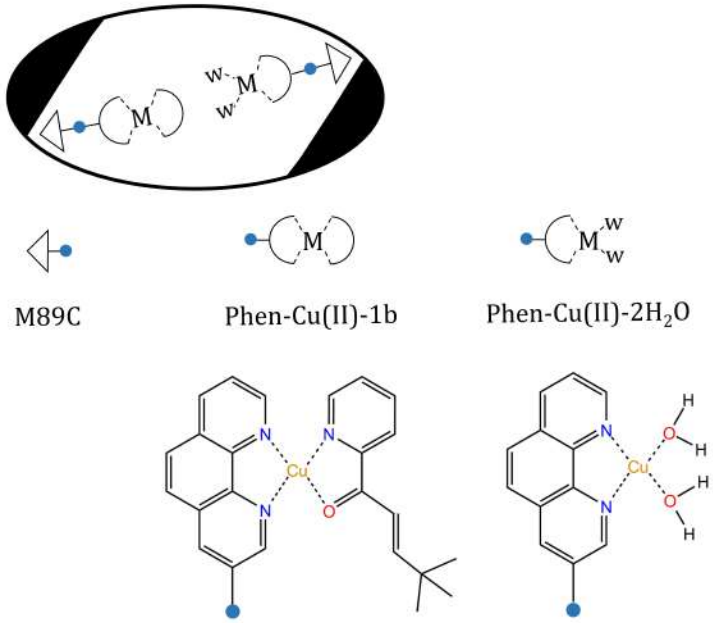


Figure 4.4: Schematic representation of the system generated after docking of Phen–Cu(II)-**2b**/Phen–Cu(II)-2H_2O at positions M89C/M89C' of LmrR, respectively.

Classical Molecular Dynamics

Force field parameters for the first coordination sphere of the metal were calculated based in the Seminario's method and RESP charges were generated with antechamber. The ions94.lib library and TIP3P [347] force fields were used for chloride ions and water, respectively. The model was set up with the xleap program [208]

The best docking solution after the second docking run was used as starting point for 100 ns MD simulation. Residues 71, 72 and 117-126 of chain A, and 1-4 and 116-126 of chain B were not determined in the X-ray structure. Residues 71 and 72 of chain A were repaired by superposition to chain B. The rest correspond to terminal

residues that were replaced by uncharged motifs proper of the AMBER force field (ACE and NME). [208]

The MD was set up with the xleap program. [208] It was immersed into a cubic box containing about 37000 water molecules (10 Å from the protein to the adge of the box) and 4 chloride counterions to neutralize the simulation cell. The TIP3P [347] and AMBER [208] force fields were used to describe water and the protein, respectively. Chloride ions were described based on the ions94.lib library.

Parameters for Phen-Cu(II)-**2b** and Phen-Cu(II)-2H_2O complexes were generated based on standard approaches. Point charges were calculated with antechamber [311] according to the RESP procedure. [305] Bonded terms involving the metal center were calculated based on the Seminario's approach. [302] The remaining atoms were parameterized according to GAFF force field. [223]

The Ewald Particle Mesh method [283] was used for the calculation of long-range electrostatic interactions, and a cut off of 10 Å was stablished for Van der Waals and short-range electrostatics interactions. The SHAKE algorithm [348] was used for constraining bonds involving hydrogen atoms. The equation of motion was integrated with a Langevin integrator [349,350] and a time step of 1 fs. A Monte Carlo barostat stablished at 1.01325 bar [351] was used to achieve constant temperature and pressure.

The simulation was run with OpenMM 7.0 program [352]. The equations of motion were integrated with a time step of 1 fs. The system was first energy minimized during 3000 steps under NVT conditions to allow progressively water molecules, side-chain and backbone to move. Then, the temperature was increased from 100 K up to 300 K for thermalization of water molecules and side chains. Last, 100 ns of production under NPT conditions were carried out. Analysis and graphical illustrations were performed with UCSF Chimera program. [291]

Quantum Mechanics/Molecular Mechanics

Pre-catalytic frames resulting from the MD simulation were selected as starting point for QM/MM calculations. Selection of frames was based on distances between the substrate double bond and the catalytic aspartate (D100'). The high layer included the Phen-Cu(II)-**2b** complex, the catalytic residue D100' and both coordinating to the metal (if any) and catalytic water molecules. This part of the system was treated at DFT level with same conditions as in previous QM calculations, excepting the Grimme's dispersion, which cannot be set up in ONIOM approach. The rest of the system was considered at MM level. Parameters for the Phen-Cu(II)-$2H_2O$ complex (included into the MM partition) were the same as the calculated for the MD simulation. All residues pointing towards the active site were allowed to move, as well as all water molecules contained into a sphere of 10 Å from the substrate **2b**.

The complexity of the system did not allow to properly polarize the Hamiltonian of the high layer by accounting for electronic embedding. Instead, mechanical embedding was the only option for the treatment of the boundaries. To overcome these problems, we went one step further and took advantage of the optimized frames resulting from the QM/MM approach to set up full QM cluster model calculations.

Full Quantum Mechanics cluster model

To assess the hydration mechanism into the LmrR protein, a full QM cluster composed of the Phen-Cu(II)-**2b** complex and all amino acids surrounding this at the active site (V15, N19, A92, F93, W96, W96', D100', I103') was performed. Quantum calculations were set up under the same conditions as the described previously, taking as starting point the optimized structures resulting from QM/MM calculations. All residues were considered up to their α carbon, that were substituted by methyl groups. With the aim to guarantee the directionality of the side chains spatial restraints on these α carbons and all their attached hydrogens were considered.

4.1.2 Results

Mechanistic study of the Phen-Cu(II) mediated hydration in water

Aiming to elucidate the two main steps of the hydration reaction (nucleophilic and electrophilic addition), a cluster model composed of the Phen-Cu(II)-**2b** complex with six surrounding water molecules was constructed. The amount of water molecules was established based on: 1) getting stabilization and proper explicit interactions with the nucleophilic water (w_N), the carboxylate group and the metal center, and 2) minimizing the drift of water molecules along the different steps of the reaction by constructing stable water chains.

Quantum calculations support the initial hypothesis suggesting a step-wise mechanism. First, the nucleophilic water (w_N) attacks the β carbon of the substrate (Figure 4.5 and Figure 4.6, $\mathbf{TS}_{N1}$). This generates an intermediate (**I1**) in which w_N preserves both hydrogens. Next, one of these protons of w_N jumps (Figure 4.5 and Figure 4.6, $\mathbf{TS}_{N2}$) to become stabilized at the water chain (**I2**). Last, the stabilized hydronium performs an electrophilic attack to the α carbon of the substrate (Figure 4.5 and Figure 4.6, $\mathbf{TS}_E$) generating the final product **P**.

The effect of the steric bulk of the R substituent of the substrate was also assessed (Figure 4.5). Results suggest that a small R group (R=methyl) easies the nucleophilic addition ($\mathbf{TS}_{N1}$) around 4 kcal/mol, in contrast to a bulky R (R=t-butyl).

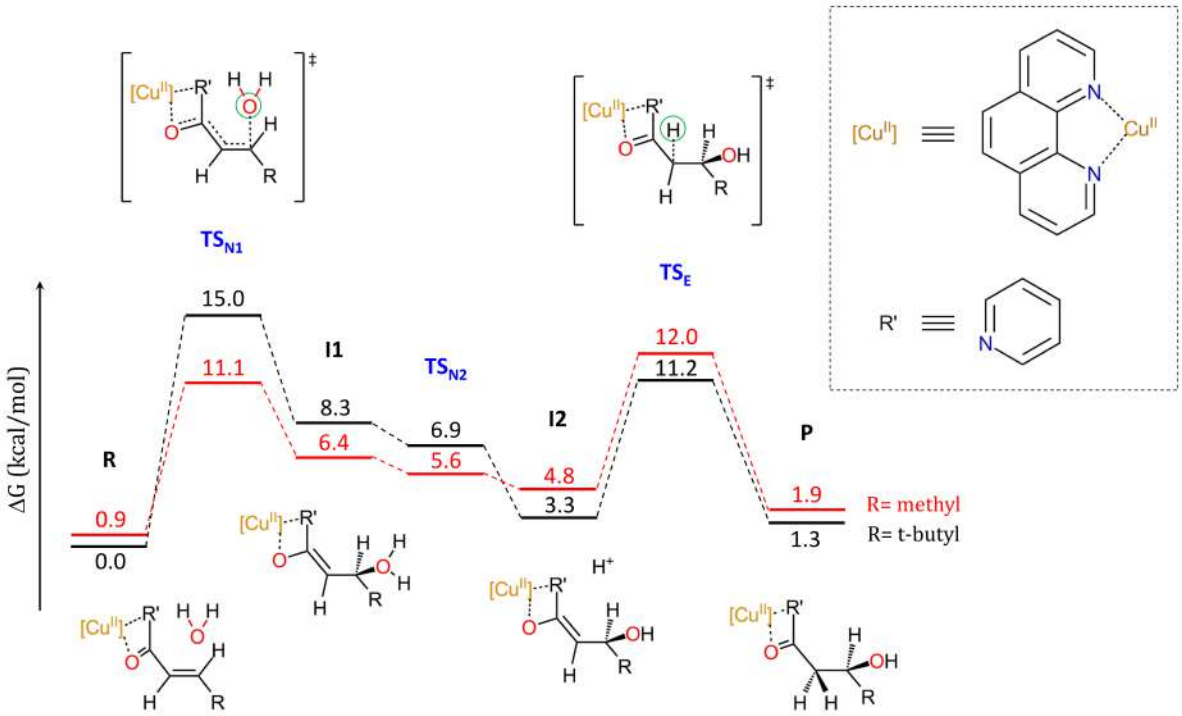


Figure 4.5: Gibbs energy profile for the copper catalysed hydration reaction in water.

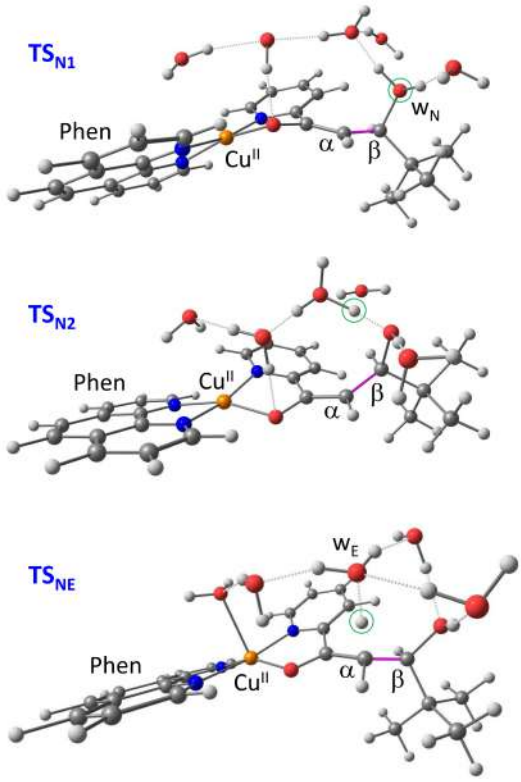


Figure 4.6: Transition state geometries for the copper catalysed hydration reaction in water. Substrate double bond and attacking atoms are highlighted in pink and green, respectively.

Aiming to discard that the suggested step-wise mechanism could be an artefact of the cluster model, we validated this point via *ab initio* Molecular Dynamics (AIMD) simulations.

From AIMD simulations we wanted to assess: 1) the nature of the water chains around the catalyst-substrate complex. Specifically, if there were water chains joining the proton of the attacking water w_N with the α carbon of the substrate. And 2) the destination of the proton after its delivery to the solvent. Regarding the first, very few real water chains were found along 12 ps of restrained AIMD simulation. They correspond with structures composed by two and four member water chains (Figure A.2). For the second, these frames together with other periodically selected, were used as starting point for 1 ps of unrestrained AIMD. After this time, the

distances between the nucleophilic water proton w_{NH} and the nucleophilic water oxygen w_{NO} were analysed to identify the delivery of the proton. In only three of the unrestrained AIMD simulations was observed the proton jumping (Figure A.3). Their visual inspection suggested that in any case the proton was transferred to the α carbon of the substrate. Instead, it appeared either stabilized in the solvent three waters away from w_N or transiting between the solvent and w_N. These results support the hypothesis that the copper mediated hydration of alkenes follows a stepwise mechanism.

To assess the diastereospecificity of the system cluster models composed of the Phen-Cu(II)-substrate (being the substrate the reaction intermediate **I**) and a pyridinium were performed (Figure 4.7). Both the S and R enantiomers of intermediate **I** as well as two different R substituents (R=methyl and R=butyl) were considered. Consistently with experimental data, results suggest that the *syn* attack ($\mathbf{TS}_E$ *syn*) is favoured upon the *anti* attack ($\mathbf{TS}_E$ *syn*) for both S and R enantiomers (Table 4.1). In this case results do not evidence a significant effect of the steric bulk of the R substituent on the stereospecificity of the reaction.

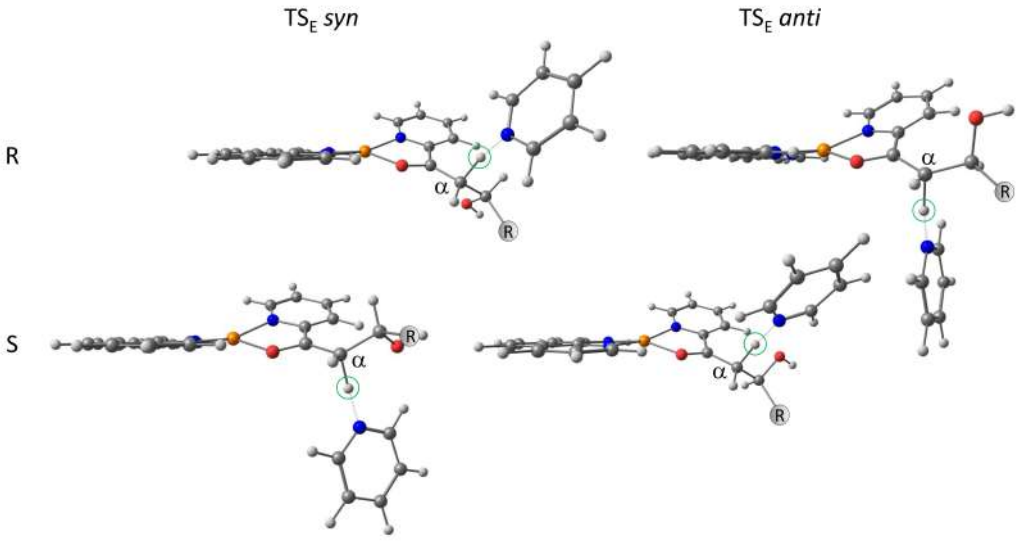


Figure 4.7: Transition state structures of the electrophilic attack mediated by a pyridinium compound over the S and R intermediates, both in *syn* and *anti*.

Table 4.1: ΔG values for the transition state structures correspoding to the electrophilic attack to the *S* and *R* enantiomers of the intermediate **I**, both in *syn* and *anti*.

	ee	**TS**$_E$ *syn*	**TS**$_E$ *anti*
R=methyl	R	0.0	3.2
	S	0.2	3.1
R=t-butyl	R	0.0	4.7
	S	0.7	4.6

Summarizing, the combination of QM calculations of the potential energy profiles and AIMD simulation supports the initial hypothesis suggesting that the copper mediated hydration of alkenes in water evolves through a step-wise mechanism. This involves: 1) a nucleophilic and 2) an electrophilic attack, being the second the rate determining step of the reaction. The first step of the reaction seems to be favoured by bulky R substituents that are directly connected to the double bond of the substrate. Regarding the second step, which defines the stereospecificity of the reaction, QM calculations support experimental data suggesting that the *syn* attack is favoured over the *anti*.

The copper mediated hydration of alkenes into the LmrR protein

The embedding of homogenous catalysts into a proteic scaffold could have different impacts. One is the modification of the mechanistic path by the intervention of second coordination sphere side chains into the reactive sphere of the metal. The second is related with binding the substrate and dictating its orientation for the reaction to occur. Before entering into the study of the full size scaffold, we decided to investigate under a cluster model approach the introduction of an aspartate moiety in the vicinity of the substrate and metal ion to mediate the activation of the water nucleophile

(Figure 4.8). This is consistent with the clear experimental impact observed about the role of D100 in the enantioselectivity of the reaction.

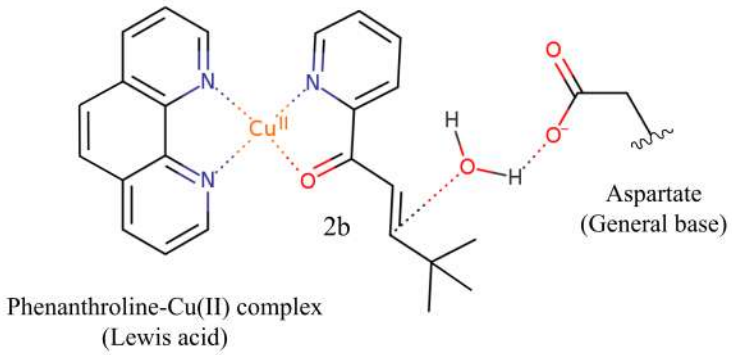


Figure 4.8: Scheme of the hypothesized water nucleophile activation mediated by an aspartate.

Two cluster models including 1) the Phen-Cu(II)-**2b** complex, 2) six water molecules (for consistency with previous model) and 3) a carboxylate moiety were constructed (Figure 4.9). Different arrangement of the water molecules, corresponding with one (Figure 4.9 a) or two (Figure 4.9 b) water molecules between the aspartate and the substrate double bond were considered.

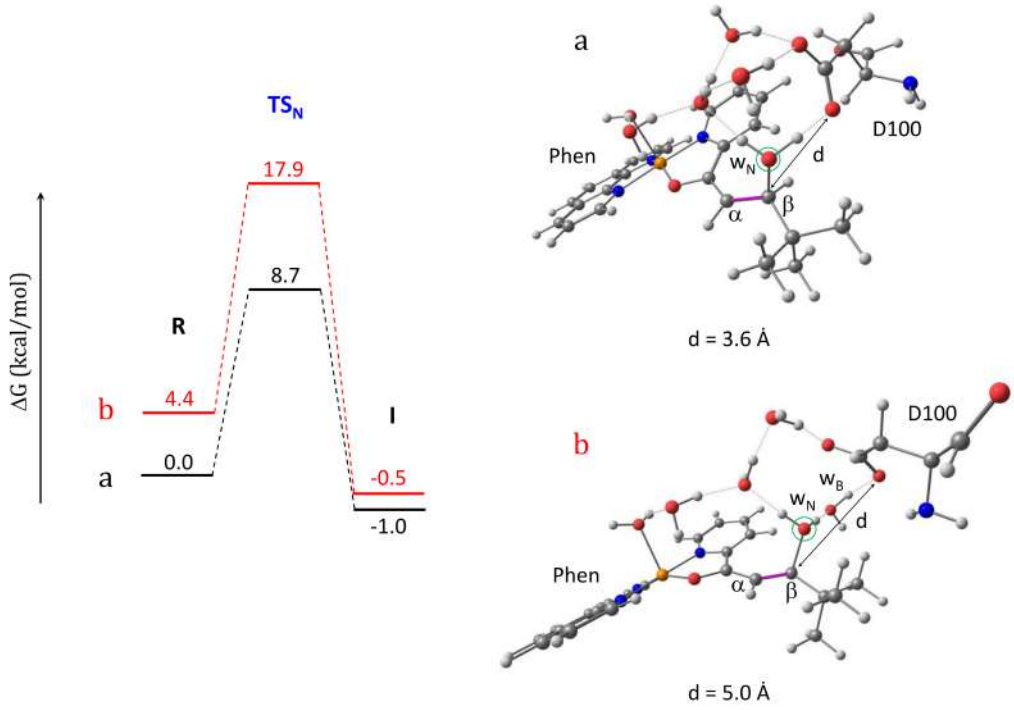


Figure 4.9: Quantum calculations for the nucleophilic attack mediated by an aspartate residue. Barriers were calculated considering either a) a direct activation of the nucleophilic water (w_N) by the aspartate or b) through a bridging water molecule (w_B).

The results suggest that the aspartate boosts the nucleophilic attack (TS_N) decreasing the barrier around 5 kcal/mol with respect to the isolated Phen-Cu(II) system (Figure 4.6). The aspartate is able to activate the water nucleophile w_N either directly (Figure 4.9 a) or through a bridging water molecule w_B (Figure 4.9 b). However, the second implies an energy penalty of around 3 kcal/mol. Additionally, results suggest that the optimum distances between the oxygens of the carboxylate and the substarte double bond is between 3.6 and 5.0 Å. This data is in agreement with structural information found in X-ray structures of natural hydratases with a bound substrate. [353–355]

Aiming to construct the full size model, Protein-ligand docking was performed. Both 1) Phen-Cu(II)-**2b** and 2) Phen-Cu(II)-2H_2O complexes were included at positions M89C and M89C' of LmrR, respectively. After the first run, which involves the

docking of the 1) Phen-Cu(II)-**2b**, the results suggest good interaction between the complex and the dimer interface (50.06 ChemScore units) (Table A.1). The complex appears sandwiched at the hydrophobic cavity between residues W97/W97' (Figure 4.10 a). Also hydrophobic interactions with residues I103', V15 and F93 were found, which seem to stabilize the complex inside the active site. Additionally, residues D100/D100' appear the only negatively charged amino acids located in the close surroundings to the substrate double bond. The second docking run involved the inclusion of complex 2) Phen-Cu(II)-$2H_2O$ into the LmrR containing complex 1 (Table A.1). In this case, results suggested a less favourable fitting (37.19 Chemscore units). The complex 2) Phen-Cu(II)-$2H_2O$ appears slightly displaced towards the solvent due to the few remaining space at the dimer interface after the inclusion of complex 1) Phen-Cu(II)-**2b**. This suggests that the LmrR-Phen-Cu(II) system would not be able to accommodate at the dimer interface two substrates at the same time.

The best docking solution resulting from the second run was submitted to 100 ns MD simulation. During this time, the Phen-Cu(II)-**2b** complex appears well stabilized at the dimer interface and shows more specific interactions with the protein active site. Residues F93/F93' gain relevance in this stabilization by performing π-stacking with both Phen-Cu(II)-**2b** and Phen-Cu(II)-$2H_2O$ complexes, respectively. The results also show stabilizing polar interactions that were not found in docking: the Phen-Cu(II) complex appears performing a hydrogen bond with N19 and interacting with D100' through the coordination of the metal center. This interaction can be direct or through a bridging water molecule located at the axial position of the copper (Figure 4.10 b). All together suggests a double role for the aspartate residue: on one hand, the stabilization of the cofactor-substrate complex into the active site and, on the other hand, the activation of the water nucleophile.

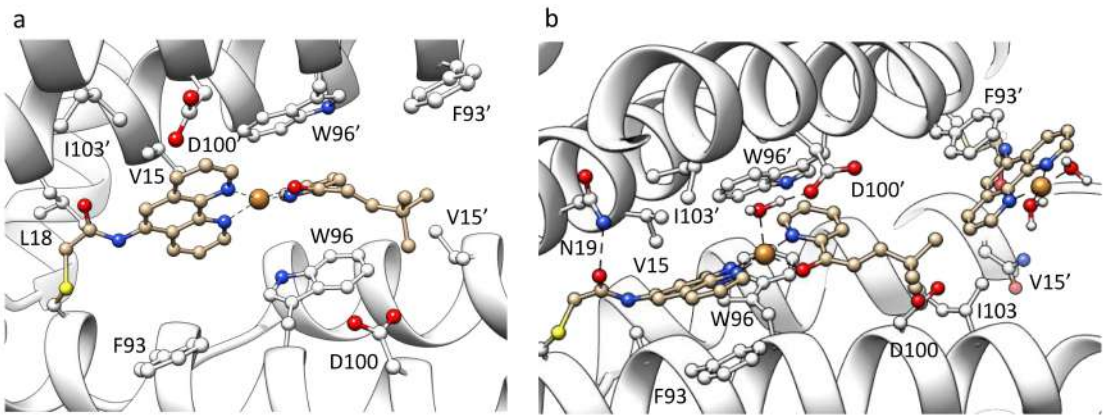


Figure 4.10: Interactions of the Phen-Cu(II)-**2b** complex at the dimer interface of the LmrR protein after a. Protein-ligand docking and b. 100 ns MD simulation.

To assess the positioning of the D100/D100' residues with respect to the substrate double bond, the distances between their carboxylate groups and the β carbon of the substrate were analysed along the trajectory. According to the cluster model (Figure 4.9), the distances found are feasible with catalysis (from 3.0 to 6.5 Å) (Figure A.4). However, visual inspection of these frames shows that only the residue D100' is able to approach the substrate double bond. In contrast, D100 appears blocked by the bulky R substituent of the substrate (Figure 4.10 b). This supports experimental observations, which relate significantly higher ee levels with the substrate **2b** (R=t-butyl) with respect to substrates with less bulky R substituents.

The water distribution around the substrate double bond was further analysed. A search of pre-catalytic configurations, i.e. frames containing water molecules at proper distances and orientations with respect to the double bond and any negatively charged residue around, was performed (Figure A.5). This allowed to determine: 1) the number of times a pre-catalytic structure is found along the simulation time scale, 2) what is the negative residue conforming such configuration and 3) at what side of the double bond the catalytic water is placed, which also allows to perform an estimation about the enantioselective tendency of the system. All pre-catalytic structures found involved the R-face of the double bond (Figure A.5) and, as expected,

D100' was the main responsible in conforming these configurations, which supports experimental data. Supporting previous observations, some of these involved the D100' directly coordinating the metal center (Figure 4.11 D100-Cu), and a others an indirect interaction through a bridging water molecule (Figure 4.11 D100-w-Cu).

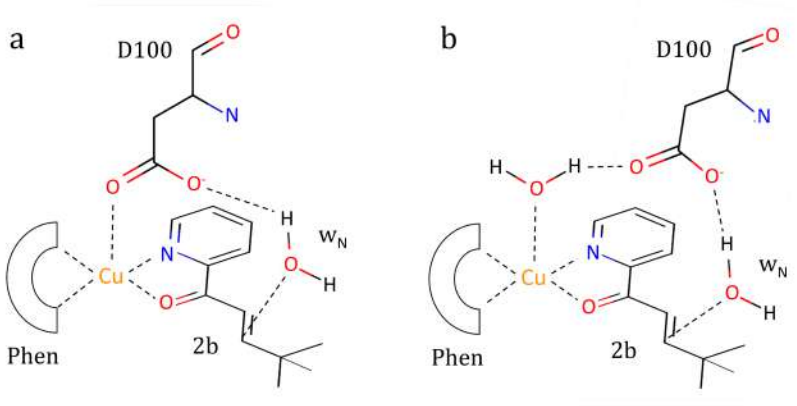


Figure 4.11: Pre-catalytic configurations found along 100 ns MD simulation, involving a. direct (D100-Cu) or b. indirect (D100-w-Cu) interaction between D100' and the metal center.

Two pre-catalytic structures involving both types of interactions between D100' and copper were selected to assess catalysis into the protein scaffold. Due to the limitations that QM/MM approaches present in this system (on one hand, leading with explicit water molecules and, on the other hand, accounting with electronic embedding) both QM/MM and full QM cluster strategies were combined to study the catalytic pathway. First, the entire system was considered at QM/MM level taking as starting point the pre-catalytic configurations. This allowed to access to the transition state geometries considering all the elements of the system, including the explicit water molecules. Next, these were reduced to models composed of less than 200 atoms for full QM analysis. These included: the Phen-Cu(II)-**2b** complex, the residues at the active site, the nucleophilic water (w_N) and the axial water coordinating the copper (if any) (Figure 4.12).

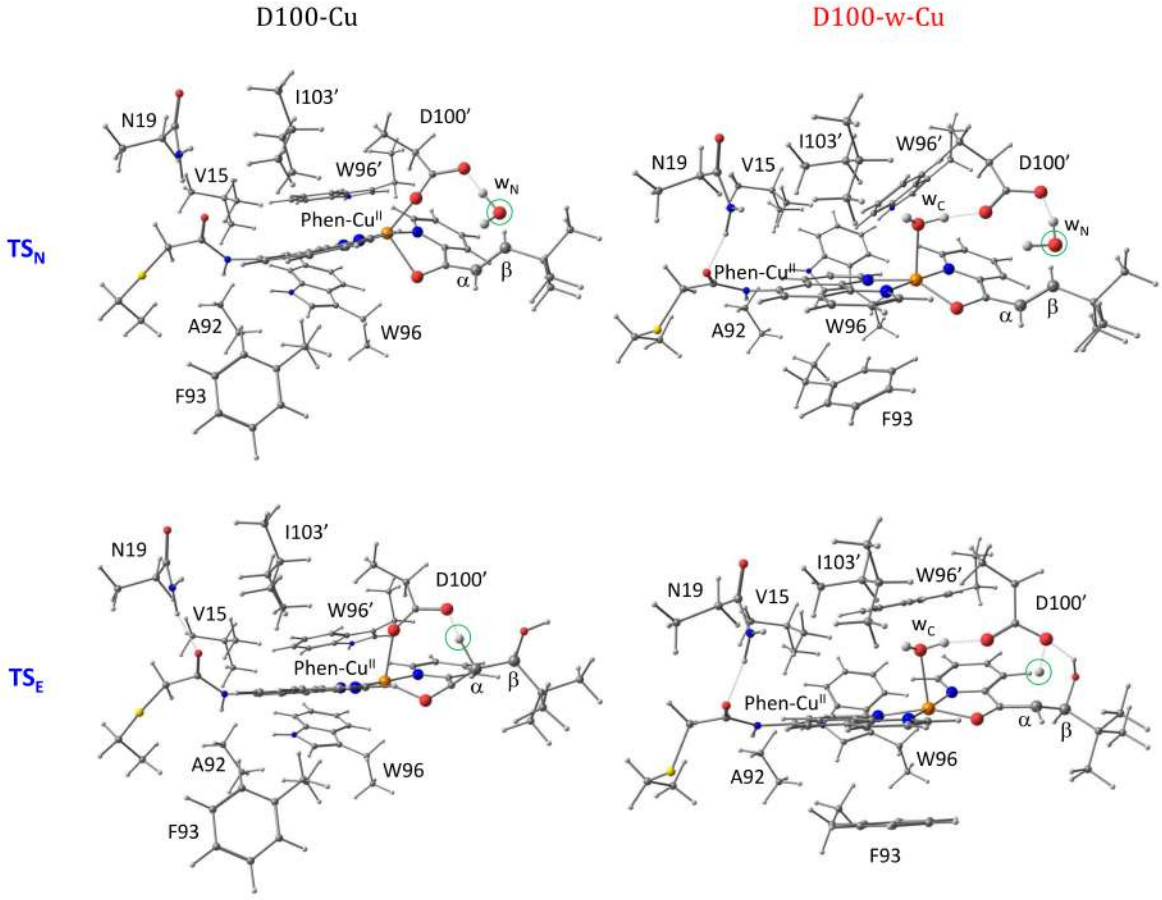


Figure 4.12: TS geometries calculated via full QM cluster models for the hydration reaction into LmrR.

Results suggest that the two steps of the hydration reaction (the nucleophilic and the electrophilic attack) can be driven by the aspartate residue (Figures 4.12 and 4.13, $\mathbf{TS}_N$ and $\mathbf{TS}_E$). For the first step, the nucleophilic water w_N attacks the double bond of the substrate at the same time that D100' extracts the remaining proton ($\mathbf{TS}_N$), leading to the intermediate **I** (Figure 4.13 **I**). This barrier (7.1 kcal/mol) is consistent with the one found in the cluster model including an aspartate moiety (Figure 4.9), and is around 8 kcal/mol less than the nucleophilic attack occurring in water. Next, the protonated D100' gives back the proton to the substrate in *syn*($\mathbf{TS}_E$) leading to the final product (**P**). These results suggest not only that the second coordination sphere favours the generation of the *R* enantiomer, but also that

the stereoselectivity is switched from *anti* to *syn*. Consistently with the reaction in water studied above, results point to the electrophilic attack (TS_E) as the rate determining step of the catalytic pathway. Additionally, the D100-w-Cu configuration appears highly favoured (around 14 kcal/mol) with respect to the D100-Cu direct coordination for both reaction stages.

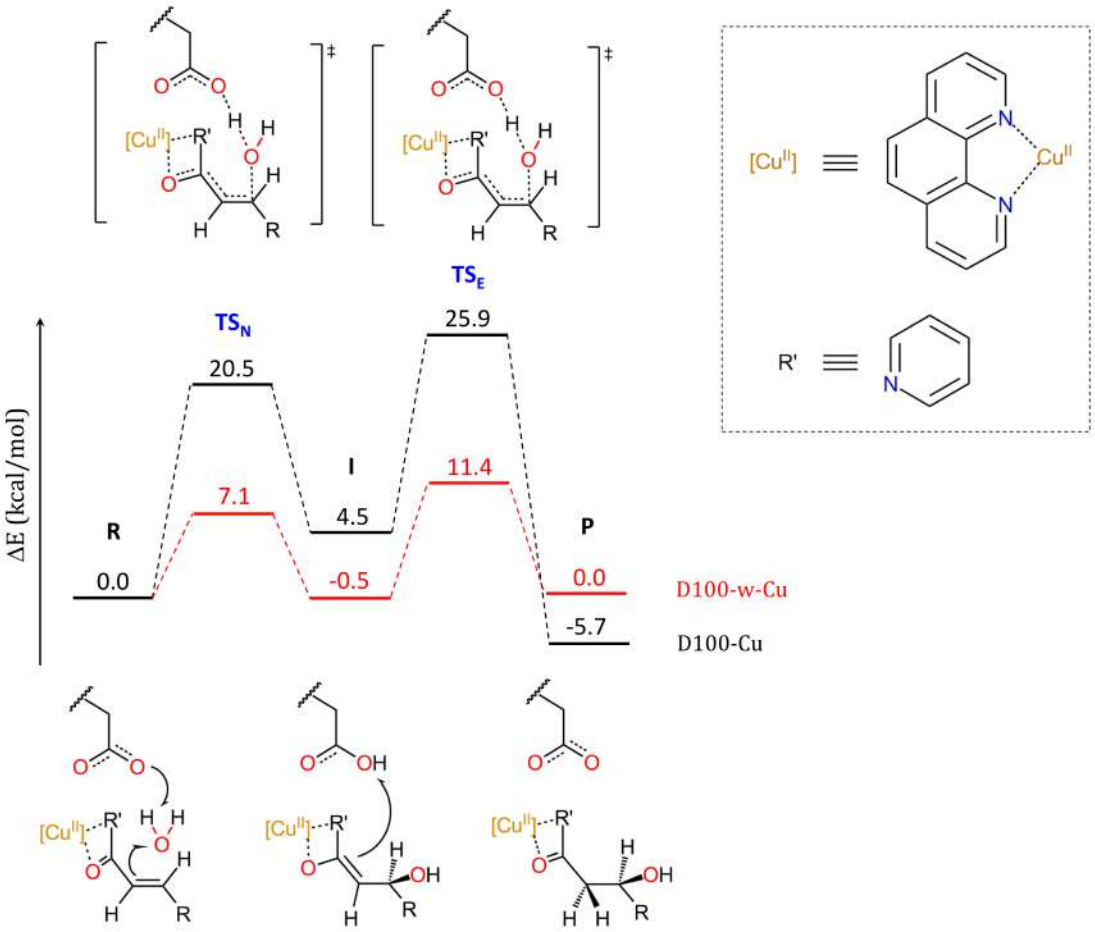


Figure 4.13: TS geometries calculated via full QM cluster models for the hydration reaction into LmrR.

These results suggest that the presence of the second coordination sphere: 1) promotes a decrease of around 8 kcal/mol for the nucleophilic attack, 2) leads to the generation of the *R* enantiomer and 3) leads the electrophilic attack to evolve in a *syn* fashion.

4.1.3 Conclusions

The copper mediated conjugated addition of water to ketones has been described based on a computational multiscale strategy. This includes from QM based strategies (such as AIMD, full QM and QM/MM calculations) to force-field based methodologies (Protein-ligand docking and MD simulations). Their proper combination has allowed to elucidate the nature of the copper mediated hydration reaction both as an isolated system and into the context of an artificial metallohydratase.

The combination of QM calculations and AIMD simulations suggest that the copper mediated hydration of alkenes in water courses through a step-wise mechanism. This involves, first, a nucleophilic and, second, an electrophilic attack, being the last the rate determining step of the reaction.

MD simulations of the full size LmrR-Phen-Cu(II) system, point to the residue D100' as the main driver of the activation of the water nucleophile. Regarding the enantioselectivity of the reaction, MD suggests that D100' approach the substrate double bond on its *R* face. Furthermore, the lack of pre-catalytic configurations involving the pro-*S* face could be substrate dependent. Bulky R substituents seems to block the accessibility of D100 to the substrate double bond. This suggestion is consistent with experimental data, which shows higher conversion but decreased ee levels for substrates with less bulky substituents. [94]

The combination of QM/MM and full QM cluster strategies allowed to elucidate the potential energy profile of the reaction into the LmrR protein. The residue D100' seems to have a double role: 1) stabilize the Phen-Cu(II) by interacting with the metal center through a bridging water molecule located at the axial position of the metal and 2) drive both the nucleophilic and the electrophilic attacks. The positioning of the aspartate with respect to the substrate double bond leads to the generation of the *R* enantiomer and the reaction to evolve in a *syn* fashion.

4.2 An artificial enantioselective imine reductase

4.2.1 An overview

The ArMs based on the biotin-streptavidin technology [356] have proven to be highly versatile and have been widely used to catalyze a wide range of new-to nature chemical reactions such as Hydrogenation, Allylic Alkilation, Suzuki Cross-Coupling and Olefin Metathesis, between others. [87]

Thomas Ward and coworkers have intensively worked on constructing new enzymes based on biotin-streptavidin technology as well as their optimization via directed evolution approaches. [357] With regard to the work of this thesis, the system of interest consists on a biotin linked [Cp*Ir(biot-p-L)Cl] cofactor included into the streptavidine (Sav) protein. This leads to an Artificial Transfer Hydrogenase (ATHase) able to perform the asymmetric imine reduction reaction (Figure 4.14). [105] This ATHase was optimized via directed evolution leading to promising 3rd and 4th generation variants for reduction of three substrates: pyrrolidine (PPL), quinolone (PDQ) or salsolidine (Figure 4.15). However, significant variations regarding both conversion and enantioselectivity were found for the different substrates. Aiming to rationalize the experimental data involving the most promising ATHase candidates, we performed a multiscale strategy which included Protein-ligand docking, metadynamics and Generalized Born Molecular Dynamics (GB MD) simulations. We aimed to describe:

- The positioning of the Ir catalyst into the Sav active site for each variant (Metadynamics)
- The assembly of the substrates with the corresponding ATHase variant (Protein-ligand docking), i.e. if they appear accommodated in a pro-S or pro-R configuration with respect to the Ir catalyst.

- The life time of the resulting assemblies from docking to discriminate between pro-S and pro-R configurations (GB MD simulations).

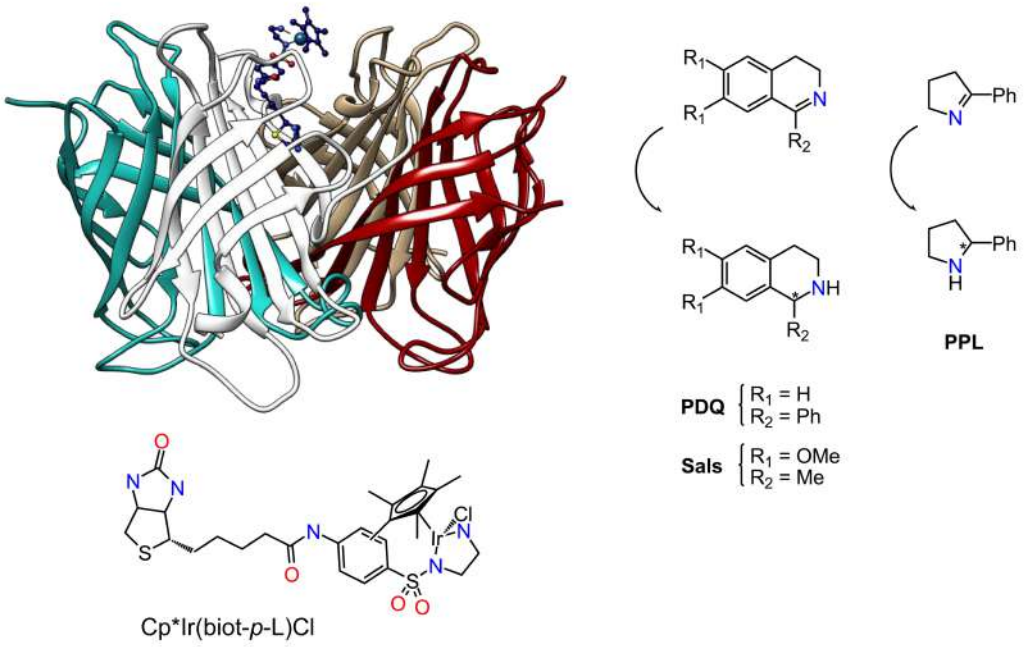


Figure 4.14: Scheme of the hydroamination reaction catalysed by the biotinylated [Cp*Ir(biot-p-L)Cl] cofactor into the Streptavidine (Sav) protein. Catalysis was tested with 1:2 equivalents of catalyst with respect to Sav monomers. i.e, two catalysts per tetramer. Each cofactor is located in an active site composed by the monomer which interacts with the biotin and the adjacent monomer. In the scheme only one cofactor is represented. The models performed in this work considers only the two monomers composing the active site for one cofactor.

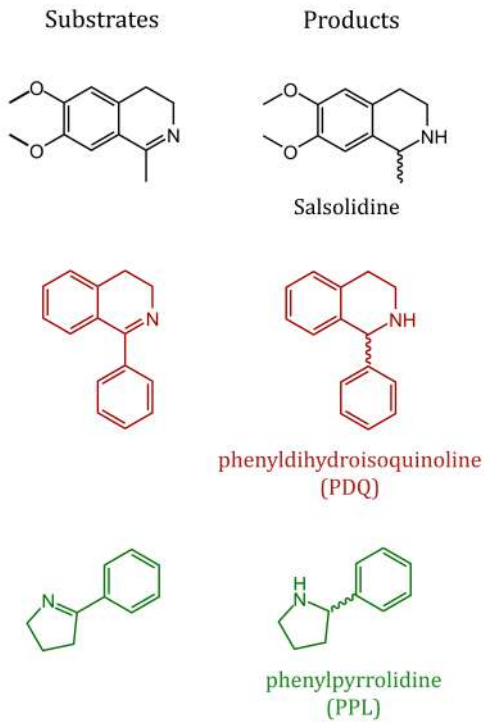


Figure 4.15: Selected substrates to assess the catalysis of the ATHase and their corresponding products.

4.2.2 The experimental input

All the experimental data related with the directed evolution experiments is collected in Figure A.6. Between the tested candidates, the most successful results were obtained for the reduction of the PPL substrate by a 3^{rd} and a 4^{th} genaration variants. Interestingly, they showed opposite enantiomeric tendencies. The 3^{rd} generation variant K121A-S112A-N118P-S122M (from now on **V1**) reduced the PPL substrate leading to an excess of 92% ee for the R product. In contrast, the 4^{th} generation variant S112R-N118P-K121A-S122M-L124Y (from now on **V2**) reduced the PPL substrate favouring the generation of the S enantiomer with a 78% ee (Table 4.2).

For salsolidine and PDQ substrates, the ee levels found were in general significantly lower. Still, variants with opposite enantiomeric tendency for both substrates were found. The most interesting case was the third generation mutant K121A-S112T-

Table 4.2: Experimental data for variants K121A-S112A-N118P-S122M (**V1**) and S112R-N118P-K121A-S122M-L124Y (**V2**) for the reduction of the PPL substrate and variant K121A-S112T-N118K-S122K (**V3**) for the reduction of PDQ and salsolidine substrates. Positive values refer to *R* ee and negative values to *S* ee.

Variants	**PPL**	**PDQ**	**Salsolidine**
V1		+36	-50
V2	+92		
V3	-78		

N118K-S122K (**V3**) which reduced the salsolidine substrate with an ee of 50% for the *S* enantiomer and the PDQ substrate with an ee of 36% favouring the generation of the *R* enantiomer (Table 4.2).

Aiming to rationalize the experimental results, the computational framework was based on comparing those systems which showed more significant variation on the enantiomeric tendency. This is, on one hand, variants **V1** and **V2** for the reduction of the PPL substrate and, on the other hand, variant **V3** for the reduction of both salsolidine and PDQ substrates.

4.2.3 Computational details

Model systems

We accounted for two X-ray structures corresponding to the variants S112A-N118P-K121A-S122M (**V1**) and S112R-N118P-K121A-S122M-L124Y (**V2**), so we focused on explaining the enantioselective tendencies for the PPL substrate taking as starting point the crystallographic material. However, these were determined assuming saturation of the Sav monomers by the cofactor while catalysis was performed under a 1:2 ratio of cofactor per Sav monomer. To properly characterize the assemblies under catalytic conditions, the dynamical behaviour of the cofactor at the protein vestibule was also assessed for these variants. Only two monomers and one catalyst

were considered for simulation. For the variant K121A-S112T-N118K-S122K (**V3**), for which no X-ray material was available, a conformation of the cofactor bound to S112A variant from a previous study (unpublished) was used as starting point. The single point mutations were generated using the Dunbrack backbone-dependent rotamer library [346, 358] contained in Chimera package. [291] Rotamers were chosen aiming to minimize the steric contacts between the cofactor and the rest of the protein.

Molecular Dynamics

The available GAFF parameters were used to describe the cofactor. Missing bonding and bending terms were generated via the Seminario's method. [302] Accordingly with a previous study, [166] the protonated form of the substrates (iminium cataions) and the hydride form of the Ir catalyst were optimized at DFT/BP86 level with the def2-TZVPP basis set using the ORCA code [359] for the generation of the force constants. The atomic charges were computed based on the RESP methodology. [305]

Implicit solvent Generalized Born (GB) Molecular Dynamics simulations were performed for the tetramer of streptavidin bound to the biotinylated Ir complex for the three variants **V1**, **V2** and **V3**. Simulations were performed with the NAMD code, [360] which implements the Born model developed by Onufriev et al., [361] at 300 K and 0 ionic strength. For the non-bonded interactions a cut off of 16 Å was used.

Models underwent 1 ns MD simulation followed by 4 ns metadynamics simulation to boost the exploration of the conformational space of the cofactor. The dihedrals around the C4 – S1 and S1 – N2 bonds were biased with a history dependent potential composed of repulsive Gaussian functions of 0.1 kcal/mol height and 10 degrees width added every 100 MD steps. During the metadynamics simulations rotation around the N1 – C1 bond also took place, so that it was not possible to assign conformational

basins to the metadynamics potential and thus, a different strategy was used. For each variant, ten conformations which showed maximal cofactor RMSD were extracted from the metadynamics trajectories and used as starting point for ten independent GB MD simulations of 2 ns long each. In most cases, the cofactor explored more than one conformation and thus, the resulting trajectories were merged and used for a clustering analysis based on the cofactor RMSD. Clusters showing a population $> 1\%$ were used to estimate the cofactor binding energy based on the GBSA/MM approach. [362]

Protein-ligand docking

Dockings were performed with GOLD5.2 program (available through the Cambridge Crystallographic Data Center (CCDC)) and evaluated based on ChemScore scoring function. [256] Aiming to discard the poses in which the enantiomeric carbon of the imine would be too far from the metal centre or blocked by the Cp star of the catalyst, distance between this carbon and the iridium were fixed to a range between 2.0 and 4.5 Å. Rotamers for all residues occupying the active site were applied using the Dunbrack library of rotamers. [346, 358]

4.2.4 Results

The asymmetric reduction of PPL by the variants V1 and V2

The positioning of the Ir catalysts into the Sav active site was assessed via metadynamics simulations. Results suggest that the catalyst is able to explore a big amount of conformations into the big active site composed by two monomers. With the aim to identify the most repeated configurations of the catalyst along the trajectory, these were grouped via clustering analysis. The most populated clusters that, in addition, were related to the lowest protein-cofactor binding energies, fitted very well with the X-ray structures for both S112A-N118P-K121A-S122M (**V1**) and S112R-N118P-

K121A-S122M-L124Y (**V2**) variants (Figure A.7). Thus, these were selected as model systems to rationalize the enantiomeric tendency for the reduction of substrate PPL.

Protein-ligand docking of the protonated form of PPL was performed. For **V1**, docking results do not show clear preference for the pro-S or pro-R binding modes. The available space around the Ir catalyst is broad enough to allow the accommodation of the substrates both in the pro-S or pro-R configurations (Figure 4.16 a). For variant **V2**, docking also suggests feasible pro-S and pro-R binding modes. In both cases, the phenyl group of the substrate lies in the pocket formed by the residues R112 and Y124 from both monomers and the residue P118 from the adjacent monomer. However, only in case of the pro-S configuration, the -NH of the substrate appeared interacting with the hydroxyl group of the Y124, while for the pro-R this was flipped by 180 degrees and did not perform such interaction (Figure 4.16 b and c). This suggest that the pro-S binding mode of substrate PPL could be favoured into variant **V2** thanks to the hydrogen bond interaction between PPL and L124Y.

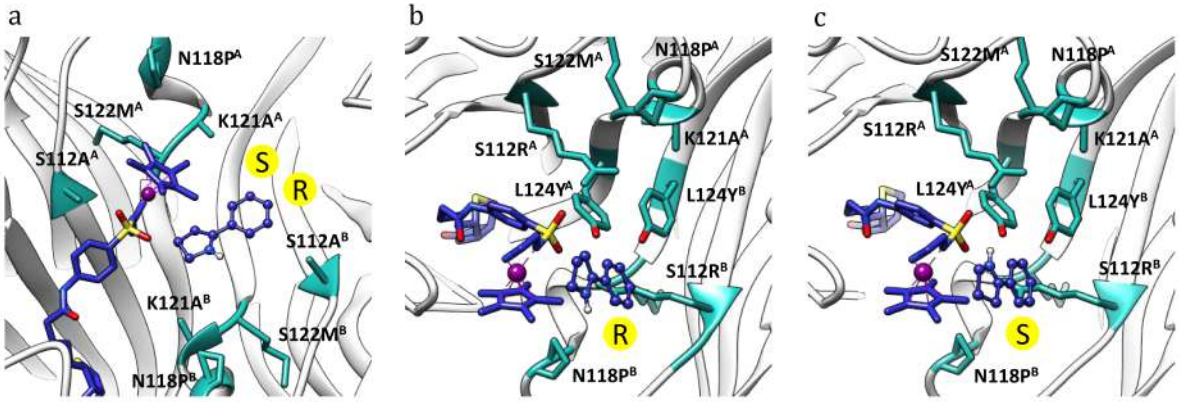


Figure 4.16: PPL binding modes into the representative structures of the most populated clusters (corresponding with the X-ray structures) for a) **V1**, which allows both pro-R and pro-S configurations (represented only the pro-R); and for b, c) **V2**, in which the pro-S binding mode (c) appears favoured by an hydrogen bond between the substrate and tyrosine L124Y^A.

To further assess the viability of the PPL binding modes at the active site of both variants, resulting complexes were energy-minimized and submitted to 2 ns of GB MD simulation.

For **V1**, the results suggest a binding mode still without clear preference for reducing either the pro-S or pro-R configuration of PPL. In both cases the substrate appears at the hydrophobic surface constituted by the residues L110, A112, T114, P118, A121 and L124. We then also considered the binding of PPL to the representative structures of other clusters related with different orientations of the Ir cofactor. However, after docking and same time scale (2 ns) MD simulations any long-lived pro-S or pro-R binding mode were found. This supports docking observations which suggest that the experimental data cannot be easily rationalized via a pure binding process for this variant. In this case, further analysis involving calculations of reaction barriers may be required.

For variant **V2**, in contrast, GB MD simulations suggested clear differences between the pro-S and pro-R binding modes. While the pro-R binding mode is not stable during the trajectory, the pro-S shows a longer life time (around 200 ps) at the nearby area of the cofactor. During this time, the pro-S binding mode appears stabilized via a hydrogen bond through its -NH group. However, this interaction is not with L124Y^A as docking suggested, but is with the carboxylate of the A121 backbone, which forms a hydrogen bond that stabilizes the binding. From this, accordingly to experimental data, we suggest that the S112R-N118P-K121A-S122M-L124Y variant should reduce preferentially the pro-S configuration of PPL, since both docking and MD simulations recognize it as the most favourable binding mode.

Assessing the enantioselective tendencies of PDQ and salsolidine into variant V3

For variant K121A-S112T-N118K-S122K (**V3**) no crystallographic material was available. Thus, as before, the most populated clusters resulting from the metadynamics which, additionally, showed best cofactor binding energies were selected as model systems. These were clusters c0, c2 and c3 (Figure A.8).

The representative structures of the selected clusters were used as host for Protein-ligand docking of the PDQ and salsolidine substrates in its protonated state. For both substrates docking found a large variety of poses related to both the pro-*S* or pro-*R* binding modes (Table 4.3). However, the representative structure of the most populated cluster c0 shows predisposition to accommodate the pro-R orientation of PDQ and the pro-*S* of salsolidine, which supports experimental data. In contrast with other variants, **V3** shows two different accessible binding sites corresponding with opposite faces of the Ir catalyst: cluster c0 binds substrates in a vestibule constituted by T112, T114, T115 and K106 residues of monomer A (Figure 4.17 a, b); instead, cluster c2 accommodates substrates next to T112, L124 and L110 residues of monomer B (Figure 4.17 b, d). With the aim to properly discriminate between the pro-*S* and pro-*R* binding modes, all the configurations resulting from docking that appeared properly oriented and at good distances from the Ir catalyst were submitted to 2 ns MD simulation.

To assess the lifetime of the selected poses next to the Ir catalyst, the MD trajectories were filtered in function of the distances between the Ir-H and the acceptor carbon of the substrates. A distance up to 4 Å was considered as it ensures pre-catalytic orientations of the substrate with respect to the Ir catalyst.

The number of extracted frames highly varied between the two substrates and also between the pro-*R* and pro-*S* binding modes. This suggests that the assessed

Table 4.3: The enantiomeric tendency of the selected poses together with the corresponding binding energies ΔG (kcal/mol) and scoring values resulting from docking.

		PDQ			**Salsolidine**		
	Cluster	Tendency	ΔG	ChemScore	Tendency	ΔG	ChemScore
V1	c0	*R*	-24.86	19.19	*S*	-18.47	13.93
		S	-24.67	19.18	*S*	-19.57	13.39
		R	-23.89	18.52	*S*	-15.30	11.97
		R	-24.86	17.84	*R*	-17.04	11.92
	c2	*R*	-32.22	19.73	*R*	-20.48	18.48
		S	-27.10	19.93	*S*	-21.19	16.75
	c3	*R*	-23.18	21.32	*R*	-16.96	15.91
		R	-21.49	19.90	*R*	-17.29	15.40
		R	-20.79	19.53			

configurations are not equally stable in the close surroundings of the Ir catalyst (Table 4.4).

For substrate PDQ, a high number of frames related to both pro-*R* (entry 1) (Figure 4.17, see PDQ c0) and pro-*S* (entry 5) (Figure 4.17, see PDQ c2) orientations were found. The pro-*S* orientation in cluster c2 appears stabilized by an hydrogen bond between the -NH of PDQ and the carboxylate group of the cofactor. However, the collected pro-*R* frames are associated to the most populated cluster c0. This suggests that, although both pro-*R* and pro-*S* orientations are stable at the different vestibules of **V1**, the system access more frequently to pro-*R* configurations, which is consistent with experimental data.

Regarding salsolidine, the MD simulations show significant preference for the pro-*S* orientation (entries 1 and 5) in the two most populated clusters c0 and c2 of **V1** (Figure 4.17, see salsolidine). As with PDQ, in cluster c2 salsolidine appears forming a hydrogen bond with the carboxylate group of the cofactor (Figure 4.17 d), which may help to stabilize salsolidine next to the metal center. In contrast, the pro-*R*

Table 4.4: Number of frames collected among 2 ns MD simulation corresponding to R or S pre-catalytic orientations between the Ir catalyst and substrates in the most populated clusters of variant **V1**.

		PDQ		**Salsolidine**		
	Cluster	Docking pose	Pre-reactive frames	Docking pose	Pre-reactive frames	Entry
V1	c0	R	280	S	162	1
		S	48	S	45	2
		R	47	S	10	3
		R	-	R	4	4
	c2	S	201	S	116	5
		R	1	R	2	6
	c3	R	284	R	18	7
		R	159	R	17	8
		R	41			9

binding modes appears unstable next to the Ir catalyst, being immediately displaced away at the very beginning of the MD simulation (entries 4 and 6-8).

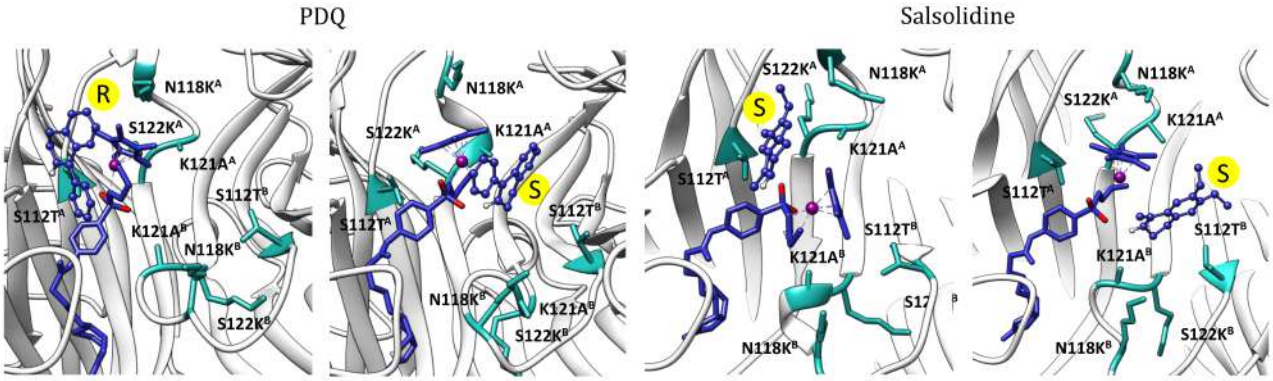


Figure 4.17: Different binding modes found for the variant **V3** for the substrates PDQ (left) and salsolidine (right).

4.2.5 Conclusions

The results of this study suggest that the enantiomeric tendencies found experimentally can be explained based on molecular recognition for three between the four assessed candidates. This are the variant **V2** reducing the PPL substrate and the variant **V3** reducing both PDQ and salsolidine substrates. The enantiomeric tendency upon reduction of PPL by the variant **V1** could not be explained basing the computational framework in the binding process. In this case, further analysis involving calculations of the energetic barriers would be required to discriminate between the S and R tendencies.

Again the results demonstrate that the proper combination of Protein-ligand docking and MD simulations is able to reproduce protein-ligand interactions that can be key to rationalize the enantioselective tendencies of ArMs. This work also supports relevance of performing post-analysis of the docking results via MD simulations. [363–367] This appears key to assess the viability of the predicted binding modes as well as to elucidate new interactions between the ligand and the protein active site.

Chapter 5

From the re-design to the *de novo* design of new functions

From the available 20 natural amino acids, the absolute sequence space include up to 20^n (being n the number of amino acids composing a protein) possibilities. The fact that the number of known proteins composing the existing organisms is of the order of 10^{12} evidences that Nature has explored only a tiny part of this space. [7]

The combination of both mutation and selection, which has driven evolution over millions of years, has spread the protein variability to cover all essential functions of living organisms. This complex process has explored a very small fraction of the available sequence space. Furthermore, protein variability has occurred in an irregular way, leading to the generation of clusters corresponding with the different protein families we now recognize (Figure 5.1).

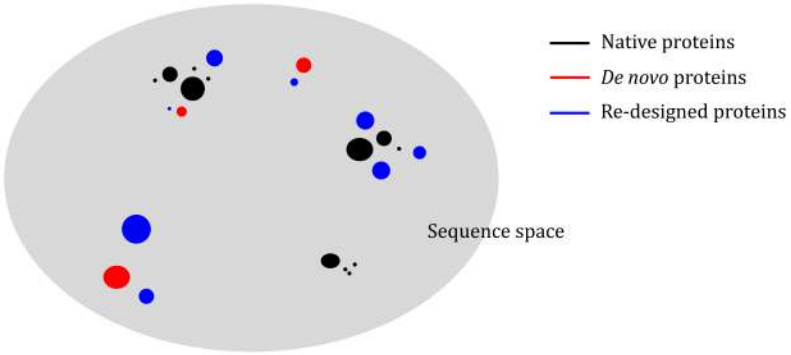


Figure 5.1: Illustration about the different ways evolution, *de novo* design and re-design explore the available sequence space. This is inspired by a recent review by Baker and coworkers. [7]

De novo design and re-design strategies aim at exploring the remaining area of the sequence space. The main difference between both is how they access to new regions. As with the *de novo* design we can displace to any point along the sequence space, via re-design we only access to close areas around pre-existing (either natural or resulting from *de novo* design) clusters.

The main difficulty for both the *de novo* design and the re-design strategies arises in the fact that, of course, not all the unexplored sequence space is able to provide functional proteins. The use of Artificial Metalloenzymes (ArM) appear as a powerful option to access to non-explored regions. First, providing a new function to a native protein scaffold (or a protein block) by incorporating an organometallic cofactor (*de novo* design). Second, optimizing this function by both rational or random changes on the sequence and/or varying the nature or the positioning of the cofactor (re-design).

In this Chapter we explore all the concepts stated above, again by making use of computational multiscale strategies. First, we perform the *in silico* study of an experimentally *de novo* designed artificial hydratase, aiming at optimizing its catalytic efficiency via rational re-design. Second, we move one step further and focus our strategy on a more challenging purpose, as is the computational *de novo* design of artificial metallopeptides, i.e. a *de novo* design with no experimental input. Here,

we demonstrate that very complex and challenging goals can be overcome with the proper use and combination of standard techniques, ranging from QM calculations to force-field based methodologies as Protein-ligand docking and MD simulations.

5.1 An artificial metallohydratase including a non-canonical amino acid

5.1.1 An overview

One of the main challenges when facing the design of ArMs is the identification of catalytically active geometries between adequate cofactor-protein partners. The use of unnatural amino acids (UAA), especially those that are able to coordinate metals, facilitates this task since it allows controlling the location of the cofactor into the protein scaffold. Roelfes and coworkers reported a new methodolgy for the *in vivo* incorporation of the metal binding non-canonical amino acid Ala-2,2'-bipyridine (from now on BpyA) into a pre-defined position of the LmrR protein [79] based on the amber stop codon suppression strategy. [131,368,369] In contrast to the covalent conjugation methodology,[1] the implementation of this approach appeared of great relevance as it allows the straightforward synthesis of the ArMs *in vivo*.

The new BpyA-Cu(II) based ArM was able to perform both the Friedel-Craft alkylation reaction [79] and the asymmetric conjugate addition of water to ketones. Despite being highly efficient for the first, this showed poor conversion and enantioselectivity for the later.

Regarding the high efficiency that the system exposed in Chapter 4 LmrR-Phen-Cu(II) shows for the hydration reaction and the high similarity between the Phen-

[1]The covalent conjugation strategy is the one implemented to synthesize the LmrR-Phen-Cu(II) artificial metalloenzyme described in Chapter 4.

Cu(II) and BpyA-Cu(II) cofactors (Figure 5.2), the low performance of the Lmr-BpyA-Cu(II) system for the hydration reaction appeared quite unexpected.

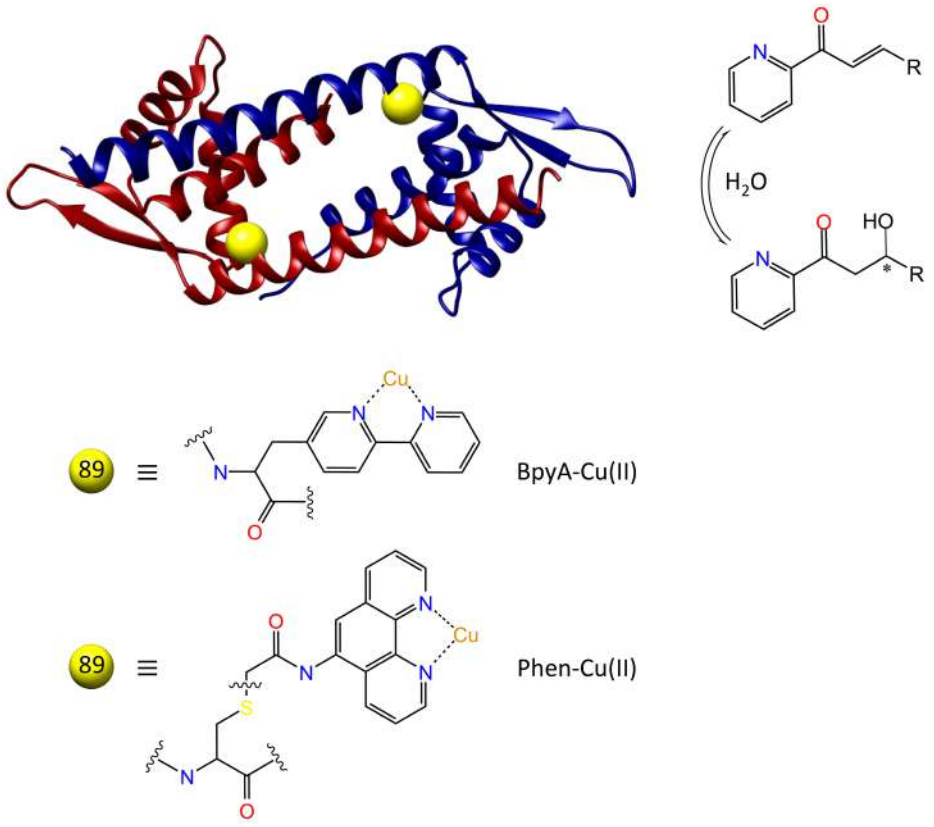


Figure 5.2: Comparison between the two copper coordinating cofactors included into LmrR to perform the conjugated addition of water to ketones. On the top, the unnatural amino acid Ala-2,2'-bipyridine (BpyA) boarded in this Chapter. On the bottom, the covalently anchored phenanthroline (Phen) cofactor, exposed in Chapter 4.

Taking advantage of the previous study about the LmrR-Phen-Cu(II) system, we decided to follow a similar multiscale strategy which would help to elucidate the main differences between this and the LmrR-BpyA-Cu(II) artificial enzyme. Once determined the key points affecting the efficiency of the BpyA based system, we proposed a variety of mutants that may improve its quality regarding both conversion and enantioselectivity.

Summarizing, we wanted to use computation, first, to study and, then, to optimize the LmrR-BpyA-Cu(II) system based on the observations resulting from the study of the LmrR-Phen-Cu(II) artificial enzyme. A similar multilevel strategy including QM, Protein-ligand docking and MD simulations was followed to characterize and optimize the second coordination sphere of the BpyA-Cu(II) catalyst (Figure 5.3). Briefly, this encompassed:

- The characterization of the activation of the water nucleophile mediated by an aspartate residue (QM cluster model) (Figure 5.3 a).
- The analysis of the receptor-ligand assembly (Protein-ligand docking) and its dynamical behaviour (MD simulations) (Figure 5.3 b).
- The rational re-design of the LmrR-BpyA-Cu(II) active site and the assessment of the new variants (Protein-ligand docking and MD simulations) (Figure 5.3 c, d).
- The experimental validation of selected candidates.

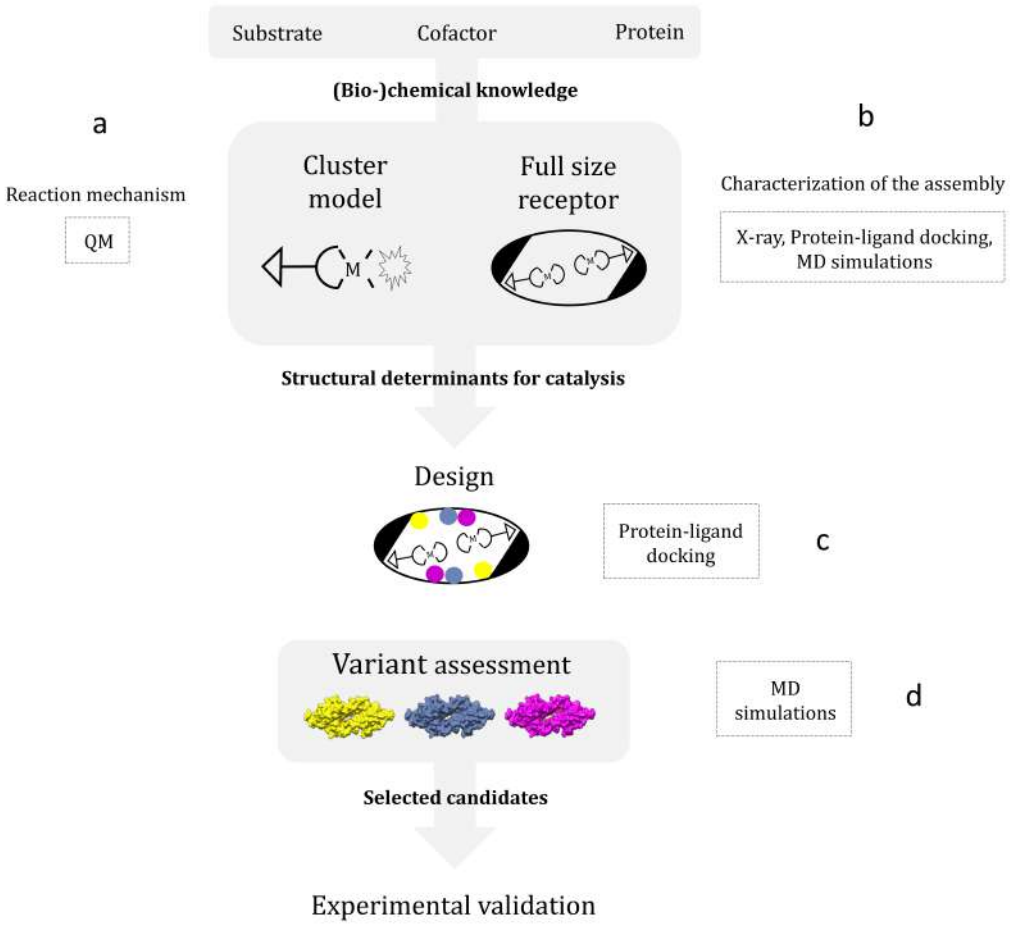


Figure 5.3: The computational framework implemented for the characterization and design of the BpyA based ArM.

5.1.2 Computational details

QM calculations

Both the optimization of the organometallic complex BpyA-Cu(II)-substrate and the cluster model calculations were performed with Gaussian 09 program [264] at DFT level using B3LYP-D3 functional. [200–202] The 6-31G(d,p) basis set [335–337] was uswed for non-metallic atoms and SDD [338] including a f polarization function for copper.

The implicit Solvation Model based on Density (SMD) was used as continuum method. [286] For those cluster models containing second coordination sphere amino acids, an epsilon adequate to describe solvent accessible surfaces of proteins (ε=9) was specified. [339]

A bi-coordinated geometry of the substrate to the copper cofactor was selected since, according to previous work, this was considered the most appropriate to fit the binding site.

In this case, the substrate **1a** (labelled according to the experimental work), [95] which corresponds to a α,β-unsaturated 2-acyl pyridine with an isopropyl as R substituent (Figure 5.4) was used as this was the one for which the first generation of the LmrR-BpyA-Cu(II) system showed the best performance.

Protein-ligand docking

Docking calculations were performed to assess the complementarity among the protein host and the cofactor-substrate complex. As there was no available structural data of the LmrR-BpyA-Cu(II) ArM, the X-ray structure of the LmrR bound to the drug Daunomycin (PDB code: 3F8F) [370] was used as receptor. Daunomycin and water molecules were removed from the model.

To link the BpyA-Cu(II)-**1a** complex at position 89/89' of LmrR, M89/M89' were mutated to A using the Dunbrack rotamer library [346] implemented in UCSF Chimera program. [291] The covalent link was imposed between the β carbon of the alanine and the respective carbon of the bipyridine ligand. To generate the homodimer two successive docking runs were carried out: the first at position M89, and the resulting structure was used to dock a second BpyA-Cu(II)-**1a** complex at position M89' (BpyA-Cu(II)-**1a**'). The resulting system will be referred from now on as LmrR_M89X (being X the BpyA-Cu(II)-**1a**/BpyA-Cu(II)-**1a**' complexes) (Figure 5.4).

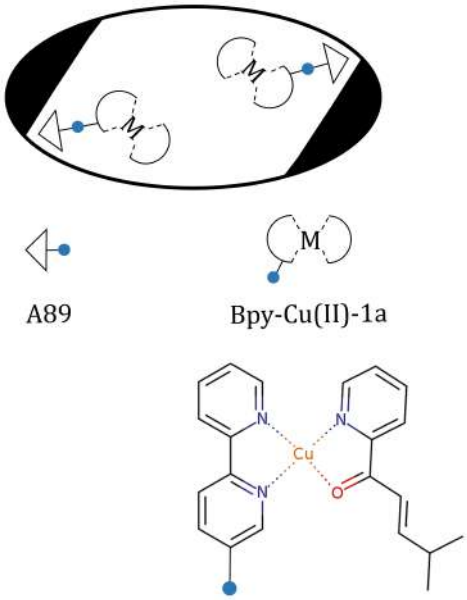


Figure 5.4: Schematic representation of the covalent linkage of the BpyA-Cu(II)-**1a** complex at positions 89/89' of the LmrR protein.

To generate the LmrR variants M89X_D100E, M89X_W96E and M89X_V15E, these second mutations were performed using again the Drunback rotamer library. [346] Then, dockings of the BpyA-Cu(II)-**1a** complex were successively performed as for the initial model.

All docking runs were carried out using GOLD 5.2 (available through the Cambridge Crystallographic Data Center (CCDC)) [236] and were evaluated with ChemScore [256] scoring function.

Molecular Dynamics simulations

The best structures resulting from docking were used as starting point to set up the models for the all-atoms molecular dynamics (MD) simulations. The side chain conformations of residues 71 and 72 of chain A, not determined in the crystal structure, were fixed by superposing the chain B. The terminal residues 117-126 of chain A and

1-4, 116-126 of chain B were not considered and uncharged terminal residues proper of the AMBER force field (NME and ACE) [208] were used to end the chains instead.

The MD simulations were set up with the xleap program. [208] The model systems were plugged into a cubic box of around 37000 water molecules and a determined number of chloride counterions (4 or 6) needed to neutralize the simulation cell. For the protein and waters, AMBER [208] and TIP3P [347] force fields were used, respectively. Chloride anions were modelled based on parameters from the ions94.lib library. [208]

Standard approaches were used to generate parameters for the BpyA-Cu(II)-**1a** complex: RESP charges [305] were calculated with antechamber. [311] The force-field terms for the description of the metal center were calculated with the Seminario's method [302] and the remaining atoms were parameterized based on the GAFF force-field [223] values.

The cutoff for the Van der Waals and short-range electrostatics interactions was set to 10 Å. The long-range electrostatic interactions were calculated via the Ewald Particle Mesh method. [283] The SHAKE algorithm [348] was used to constrain bonds engaging hydrogen atoms. The equation of motion was integrated with a Langevin integrator [349, 350] and using a time step of 1 fs. To get constant temperature and pressure the system was coupled to a Monte Carlo barostat at 1.01325 bar. [351]

OpenMM 7.0 program [352] was used as engine to run the MD simulations. A 3000 steps energy minimization was initially performed to progressively allow accommodating the water molecules, side chains and the backbone; then, the temperature was increased from 100 to 330 K to allow thermalization of water molecules and side chains; after that, 100-150 ns (depending on the stability of the systems) of production under NPT conditions were carried out and further analysed.

Molecular visualization and graphics were performed with the UCSF Chimera program. [291]

5.1.3 Modelling the first generation of the artificial enzyme

Assessment of the water nucleophile activation: a cluster model

Consistently with the previous study about the LmrR-Phen-Cu(II) system, a cluster model composed of: 1) the BpyA-Cu(II) bound to the substrate **1a** in a bidentate fashion and with a square planar geometry, 2) six surrounding water molecules and 3) a carboxylate moiety able to activate the water nucleophile was constructed (Figure 5.5).

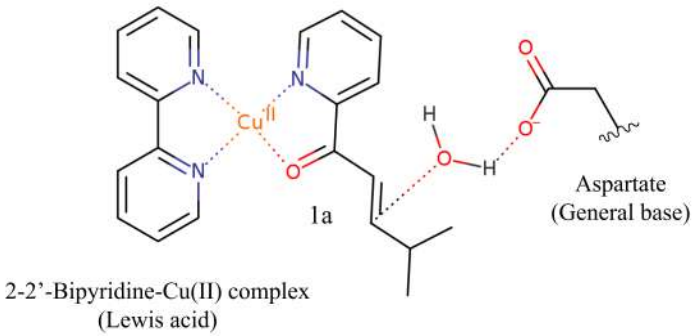


Figure 5.5: Water nucleophile activation strategy for the enantioselective hydration of alkenes into the LmrR protein. The 2,2'-bipyridine is the side chain of the BpyA unnatural amino acid bound to positions 89 and 89' at the LmrR scaffold.

QM calculations of the nucleophillic activation (TS_N) revealed both the proper arrangement and a low activation barrier (8.7 kcal/mol) for the addition of water to the conjugated double bond of the ketone (Figure 5.6). Calculations with either one or two bridging water molecules between the oxygen of the carboxylate and the β carbon of the attacked double bond suggest that the attack mediated by a direct interaction between the aspartate and the nucleophillic water (w_N) (Figure 5.6 a) takes place with a barrier around 5 kcal/mol lower than if this activation occurs through a bridging water molecule (w_B) (Figure 5.6 b). Results also suggest that the required distance between the carboxylate and the β carbon of the substrate lies into a range of about 3.5 to 5.0 Å. This data is in agreement with the results obtained for

the Phen-Cu(II)-**2b** system in Chapter 4, which at the same time is consistent with available structural data of hydratases with a bound substrate. [353–355]

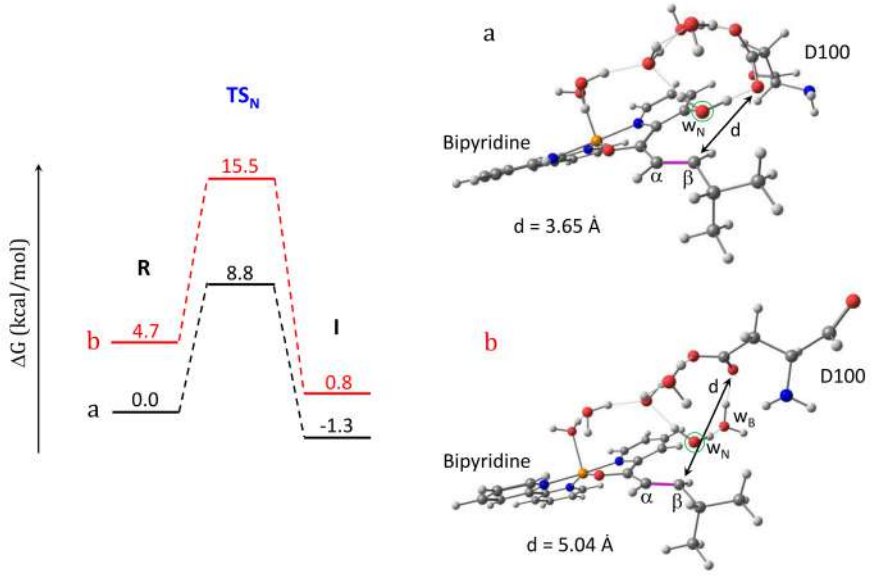


Figure 5.6: QM cluster model to assess the copper mediated hydration of ketones. The model is composed by the substrate **1a**, six explicit water molecules and an environmental aspartate which acts as a Lewis acidic base; in a) the nucleophilic water w_N donates its H^+ directly to the aspartate, while in b) a bridging water w_B mediates the transference of the H^+ between w_N and the base.

Full model of the LmrR-BpyA-Cu(II)-substrate triad: Protein-ligand docking and MD simulations

Aiming to generate 3D models of the LmrR bound to the BpyA-Cu(II)-**1a** complex (further referred to as LmrR_M89X) the structure of the BpyA-Cu(II)-**1a** was optimized via QM calculations. Next, covalent Protein-ligand docking was performed to include it at the position 89 of each monomer of the LmrR scaffold.

These simulations suggest that, in contrast to the LmrR-Phen-Cu(II) system, the LmrR-BpyA-Cu(II) artificial enzyme is able to accommodate two substrates at the same time at the dimer interface. This is reasonable since the tail connecting the bipyridine to the α carbon of the backbone is four atoms shorter. This

also implies worst interaction energies between the BpyA-**1a** complex and the dimer interface (33.76 ChemScore units) with respect to the LmrR-Phen-Cu(II) system (50.06 ChemScore units), as the former is not able to properly approach the center of the hydrophobic cavity (Table B.1). Still, both BpyA-Cu(II)-**1a**/BpyA-Cu(II)-**1a**' complexes show good hydrophobic complementarity with surrounding hydrophobic residues (Figure 5.7 a). Edging or π-stacking interactions between the aromatic rings of both bipyridines and F93/F93' are also observed. In addition, the pyridine of both substrates appears making hydrophobic interactions with V15/V15', while the isopropyl groups are packed between the tryptophans W96/W96'. As for the LmrR-Phen-Cu(II) system, residues D100/D100' appear the only potential general bases in the close surroundings of the substrates.

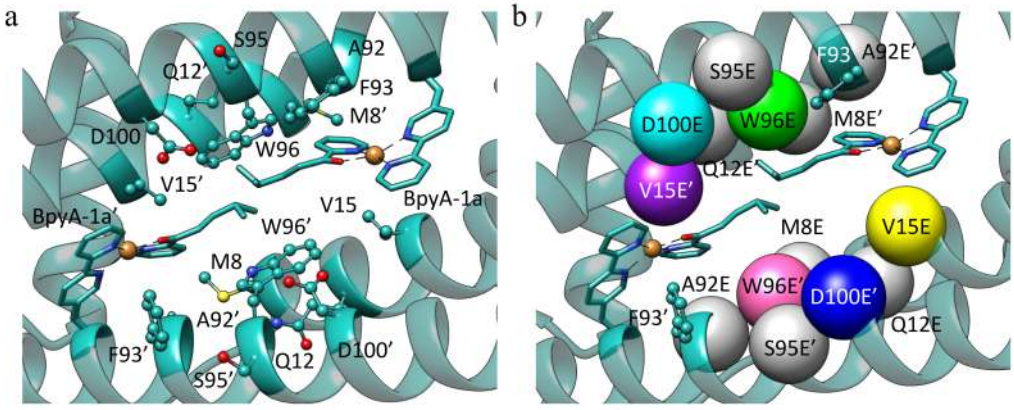


Figure 5.7: a) Docking of BpyA–Cu(II)-**1a**/BpyA–Cu(II)-**1a**' at position 89/89' of LmrR. b) Residues selected for the introduction of the glutamate that serves as general base. The models including mutations coloured in grey were not considered as they resulted in non-reactive arrangements of the protein-cofactor-substrate assembly.

Taking as starting point the best docking solution, 100 ns MD simulation were generated to explore the dynamical behavior of the new system. Its analysis revealed the following features:

- Both BpyA-Cu(II)-**1a** complexes remain well stabilized at the dimer interface mainly by hydrophobic interactions with F93/F93' and W96/W96'. However,

results reveal certain flexibility in such a way one of the cofactors is occasionally displaced towards the entrance of the active site becoming more exposed to solvent.

- The linker connecting the bipyridine to the backbone is not long enough for the copper to coordinate the D100/D100' residues (as was observed with the LmrR-Phen-Cu(II) system). This also makes the BpyA-Cu-**1a** complexes to visit the entrance of the cavity more frequently than Phen-Cu(II)-**2b**, which could be one of the factors affecting the enantioselective capacity of the LmrR-BpyA-Cu(II) system.

- The presence of the second coordination sphere is suggested to induce enantioselectivity since the water accessibility to the prochiral faces of the substrate is not equivalent (Figure B.1, first row).

- The distances between the oxygens of the D100/D100' (the only candidates to activate the water nucleophile) and the β carbon of the double bond of the substrates are generally too large along the majority of the MD trajectory. However, pre-reactive configurations (those in which the catalytic water is properly located and oriented with respect to the carboxylates and the double bond of the substrate) involving D100/D100' are found in about a 10% of the MD frames (Table B.3, M89X). In addition, most of these configurations involve the pro-R face of the substrate. All this suggests that the first generation of the LmrR_M89X enzyme should have some catalytic activity and preferentially form the R enantiomer.

5.1.4 Optimization stage via rational re-design

The observed changes on both the nature of interactions (displaced cofactors towards the solvent) and the sub-optimal carboxylates-substrate distances seem to negatively

affect the catalytic profile but also reveal a way for optimization. Thus, docking results were further analysed to identify key amino acids that could be mutated based on our structural criteria for an efficient hydration. This was focused on 1) stabilizing the cofactor-substrate complex at the dimer interface for preventing solvent exposure and/or 2) decreasing the distances between the substrates double bonds and a general base (aspartate or glutamate). For that purpose, all the residues contained into a sphere of 7 Å from the substrates (considering previous docking to the WT LmrR protein) were mutated to glutamate (Figure 5.7 and new rounds of Protein-ligand docking of the BpyA-Cu(II)-**1a** in all of the resulting variants were performed.

Only D100E, V15E, W96E and F93E were found to improve the distances form their terminal side chain atoms and the electrophilic carbon of the substrate **1a** (Table B.4). However, F93E mutation was not considered due to the observed stabilizing interactions between its aromatic ring and the cofactor-substrate complex. Best docking solutions on the mutated residues 8, 12, 92 and 95 to glutamate were characterized by too long distances between the carboxylates and the double bond (Table B.4) and thus, these were neither considered for further analysis.

Hence, best docking solutions for LmrR_M89X_D100E, LmrR_M89X_W96E and LmrR_M89X_V15E mutants were selected as starting points for 100 ns MD simulations.

In case of LmrR_M89X_D100E system, the dynamical behaviour of the two cofactor–substrate complexes is not equivalent during the simulation. Intuitively, it was expected that mutagenesis of D100 to E may bring the general base closer to the cofactor-substrate complex, resulting in improved activity. However, the MD simulation suggests that such an improvement is not performed. BpyA–Cu(II)-**1a** explores three different equally populated conformations, one of which features the substrate outside the cavity. Pre-reactive conformations slightly favouring the pro-*R* face of the double bond are found involving E100, E100' and E107 residues (Figure 6.4 b

and Table B.2, M89X_D100E). For BpyA–Cu(II)-**1a**' complex, two conformations are observed along the time-scale of the simulation and a small set of pro-R pre-reactive conformations are identified involving E100 residue (Table B.2, M89X_D100E). These results suggest that LmrR_M89X_D100E mutant should perform reactivity and enantiomeric tendencies similar to LmrR_M89X.

Along the MD simulation of LmrR_M89X_W96E, both cofactor-substrate complexes display a similar behaviour staying predominantly at the dimer interface (Figure 6.4 c). Interestingly, their orientation significantly varies with respect to the previous systems allowing different residues to approach the substrate double bond. This results in a higher number of pre-reactive configurations involving more pro-S than pro-R faces. E97 and E97' are found to interact with both substrates and D100 and D100' appear also the main responsible in activating the water nucleophile (Figure 6.4 c, B.2 and Table B.2, M89X_W96E). These data suggests that LmrR_M89X_W96E mutant should display higher catalytic activity than the first generation system but with opposite enantiomeric tendencies.

In the case of LmrR_M89X_V15E, the BpyA–Cu(II)-**1a** complex appears more exposed to the solvent than BpyA–Cu(II)-**1a**'. BpyA-Cu(II)-**1a** is generally disposed perpendicular to the axis of LmrR. In this conformation, E15 is able to perform pre-reactive configurations involving the pro-R face three times more often than the first generation mutant (Table B.2, M89X and M89X_V15E). Additionally, fluctuations of the cofactor substrate complex allow D100' to also activate the water nucleophile even more frequently than E15 (Figure 6.4 d, Table B.2, M89X_V15E). On the other hand, BpyA–Cu(II)-**1a**' maintains only one conformation along the simulation due to the interaction between E15' and the metal center (Figure 6.4 f). In this stabilizing configuration it is also able to approach the pro-R face of the substrate, suggesting a double role for E15': stabilization of the complex inside the cavity and activation

of the water nucleophile in a pro-R fashion. These results suggest that this mutant should display higher activity and same enantiomeric tendency than LmrR_M89X.

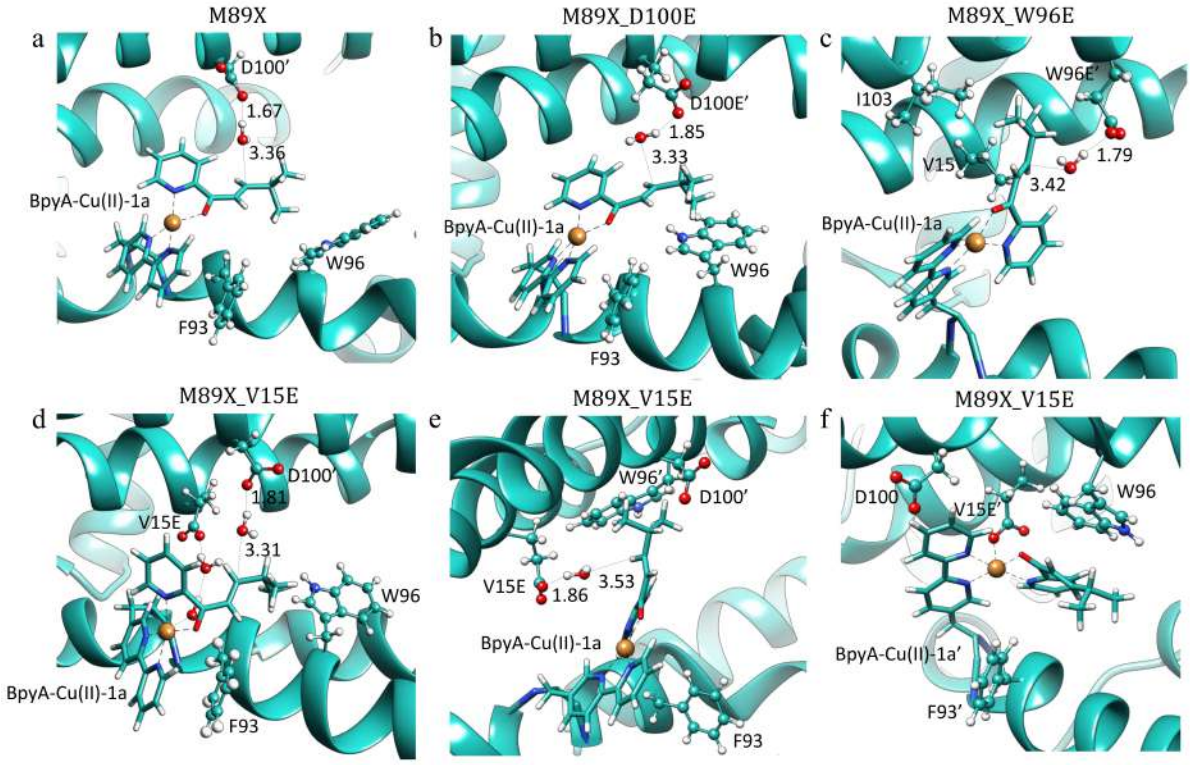


Figure 5.8: Representative pre-reactive conformations from 100 ns molecular dynamics simulations of a) LmrR_M89X, b) LmrR_M89X_D100E, c) LmrR_M89X_W96E and d-f) LmrR_M89X_V15E. These configurations are found in 10% of the simulation for LmrR_M89X, 10% for LmrR_M89X_D100E, 91% for LmrR_M89X_W96E and 31% for LmrR_M89X_V15E (Table B.2)

.

5.1.5 Experimental validation

The activity of the three designed enzymes was tested to yield the corresponding β-hydroxy ketone product **2a** (Table 5.1). Experimental results demonstrate that the three systems gave rise to an increase of the conversion of substrate **1a**.

The mutant LmrR_M89X_V15E is the most interesting case, which displays an expected increase in both conversion and enantioselectivity (Table 5.1, entry 7). In the case of LmrR_M89X_D100E, experimental results support computational predic-

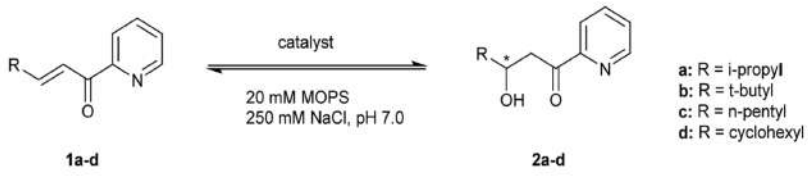
tions about the inefficiency of this mutation since only slightly better activity is found and the enantioselectivity decreases with respect to the first generation system. As expected for the LmrR_M89X_W96E mutant, it provides the highest increase in conversion accompanied with a decrease on entioselectivity (Table 5.1, entry 9). However, the enantiomeric preference was the opposite with respect to LmrR_M89X, which contradicts computational predictions. The fact that this mutant experimentally gives rise near racemic product should be due to the high contribution of different native residues to catalysis (via computation were observed 12 residues able to constitute pre-reactive configurations), which makes the computational prediction more difficult. Additionally, it should be considered that obtaining near racemic product means that this mutant presents stronger preference for the formation of the opposite enantiomer than the other variants.

The relevance of properly placing the general base with respect to the Cu(II) and the substrate double bond was confirmed by additional mutagenesis introducing glutamate to glutamine, an amino acid similar in size but lacking the negative charge. The LmrR_M89_D100Q and LmrR_M89X_V15Q mutants display similar activity and experience a significant decrease of both activity and enantioselectivity, evidencing the role of the glutamate for the selective placement of the water nucleophile in a prochiral face of the substrate. In contrast, the activity of the LmrR_M89X_W96Q mutant is similar to the W96E system, supporting the computational observations about the high amount of residues able to perform pre-reactive configurations.

5.1.6 Conclusions

We designed a new artificial metallohydratase enzyme containing an UAA (BpyA) able to catalyse a chemically challenging reaction: the selective conjugate addition of water to alkenes. Its design has been based on combining intuition and both structural (the generation of a stable metal binding site) and (bio)chemical (the mechanism of

Table 5.1: Experimental results of the enantioselective hydration reaction performed by the LmrR_M89X and its variants. [95]



1a-d 2a-d

Entry	Catalyst	Substrate	Product	Conversion (%)	ee (%)
1	—	**1a**	**2a**	11 ± 3	—
2	$Cu(NO_3)_2$	**1a**	**2a**	83 ± 9	—
3	LmrR_M89X (no Cu(II))	**1a**	**2a**	9 ± 3	—
4	LmrR_M89X_Cu(II)	**1a**	**2a**	39 ± 7	42 ± 6
5	LmrR_M89X_D100E_Cu(II)	**1a**	**2a**	50 ± 6	30 ± 1
6	LmrR_M89X_D100Q_Cu(II)	**1a**	**2a**	35 ± 3	<5
7	LmrR_M89X_V15E_Cu(II)	**1a**	**2a**	75 ± 9	64 ± 2
8	LmrR_M89X_V15Q_Cu(II)	**1a**	**2a**	28 ± 5	15 ± 3
9	LmrR_M89X_W96E_Cu(II)	**1a**	**2a**	79 ± 3	6 ± 4
10	LmrR_M89X_W96Q_Cu(II)	**1a**	**2a**	64 ± 1	6 ± 1
Substrate scope[b]					
11	$Cu(NO_3)_2$	**1b**	**2b**	86 ± 4	—
12	LmrR_M89X_Cu(II)	**1b**	**2b**	37 ± 2	41 ± 4
13	LmrR_M89X_V15E _Cu(II)	**1b**	**2b**	64 ± 8	50 ± 2
14	$Cu(NO_3)_2$	**1c**	**2c**	26 ± 4	—
15	LmrR_M89X_Cu(II)	**1c**	**2c**	24 ± 4	19 ± 5
16	LmrR_M89X_V15E_Cu(II)	**1c**	**2c**	58 ± 5	14 ± 1
17	$Cu(NO_3)_2$	**1d**	**2d**	37 ± 12	—
18	LmrR_M89X_Cu(II)	**1d**	**2d**	17 ± 4	22 ± 2
19	LmrR_M89X_V15E_Cu(II)	**1d**	**2d**	45 ± 5	57 ± 3

[a] Standard conditions: 9 mol% $Cu(H_2O)_6(NO_3)_2$ (90 μM) loading with 1.25 eq. LmrR_X (in monomer) in 20 mM MOPS buffer (pH 7.0), 250 mM NaCl, for 3 days at 4 °C. All data are the average of 2 independent experiments, each carried out in duplicate. Errors are reported as standard deviation.
[b] Conditions the same as in the experiments with **1a**.

the reaction) knowledge on natural and artificial enzymes with *in silico* modelling protocols.

Starting from a set of empirical conditions, an initial hypothesis was gained by combining QM, docking and MD simulations for the modelling of the LmrR_M89X template. Next, *in silico* analysis drove the optimization of the second coordination sphere of the metal for the adequate placement of a general base at the designed active site. The positions proposed were never considered before on the previous empirical design, [79] evidencing the relevance of using *in silico* approaches especially when lacking structural information of the artificial enzyme. The study of the new

variants through the combination of docking and MD simulation approaches allowed: 1) qualitative predictions about the activity and enantiomeric tendencies of the reaction, the last based on comparisons of the prevalence of pro-S or pro-R pre-reactive conformations for the different mutants; and 2) suggesting interactions that were not previously anticipated, such as the one between the Cu(II) ion and the introduced glutamate residues in the mutant LmrR_M89X_V15E.

Results demonstrated that computation is able to provide suitable models of the LmrR system that, in general, were in good agreement with the experimental catalytic results. This supports other works in which computation resulted a powerful tool for the optimization of new ArMs and to elucidate their catalytic mechanisms. [79, 323, 370–372] Experimental verification demonstrated that the combination of QM, Protein-ligand docking and MD simulations is an adequate approach for the design and optimization of artificial metalloenzymes. These are the first set of *in silico* rational designed enantioselective artificial metalloenzymes based on a catalytic UAA so far.

5.2 Computational *de novo* design of artificial metallopeptides

5.2.1 An overview

Predicting the folding of small proteins and peptides has become one of the bigest challenges in structural biology and protein design fields. Their study has gained relevance since the discovery of active peptides involved in relevant biological processes and neurodegenerative diseases, [373] which has converted them in excellent targets for therapeutic applications. [374] Designing peptides for therapeutic purposes requires achieving structures of high stability, diverse functionality and favorable phar-

macokinetic properties. In this context, macrocyclic peptides [375–378] and the so called hyper stable constrained peptides [24] appear as a promising starting point for designing new functions.

Incorporating metal moieties at biological scaffolds leading to metallopeptides has been of great relevance for elucidaing how they fold and/or unfold, aggregate [379] or interact with other biological structures. [380] Probably, the most relevant studies in recent years involving metallopeptides are related with the role of metal ions in the generation of amyloid fibrils in Alzheimer disease. [379,381–384] They have been also successfully applied to the study of protein-protein interaction processes of different nature [385] and to asymmetric catalysis. [386] More specifically, important reactions such as the enantioselective oxidation of indoles [387] and the asymmetric sulfonylation, [388] between others, [386] have been approached.

Many efforts have been dedicated to re-engineer naturally occurring peptides (especially stable constrained peptides) based on loop grafting, selection and sequence randomization techniques to generate new bioactive molecules. [389] Despite being powerful, these approaches are limited by the small amount of these naturally occurring structures. Furthermore, the challenging control on the size and shape of the new molecules makes it difficult to find global complementarity with their targets. In this context, the use of computational approaches to drive the design of stable structures appears as a promising challenge that may allow gaining control on the structural behavior of the new constructs. G. Bhardwaj and coworkers have been recently demonstrated the power of computation for the *de novo* design of new peptides based on generating initial robust scaffolds followed by experimental selection approaches. [24] In this context, Rossetta [26, 27] program gains relevance due to its successful contribution to predict active scaffolds from scratch. The use of stabilizing moieties such as N-C cyclized backbones, disulfide constrains and non-canonical side chains appear promissing for achieving new accurate designs. [390] Among the

reported successful metallopeptides driven by computation, their design have been based mainly on including multiple metal binding sites, generating complex tertiary structures and the assembly of polypeptide blocks. [391–397] Despite the impact of these studies, the link between the structure and the catalytic functions continue in its infancy and any contribution to identify structural keys to get stable and functional tinny scaffolds is welcome.

Here, we have taken advantage of the chemical and structural knowledge acquired along the previous studies included in this thesis, and present a set of *de novo* designed peptides driven by *in silico* approaches to perform the asymmetric conjugate addition of water to ketones. Regarding the different stages for the *de novo* design of ArMs exposed in Chapter 1, we referred to the **discovery** stage as the part of the design dedicated to find new enzyme-cofactor partners with a new functionality; and to the **optimization** stage as the required modifications for the new ArMs to become efficient catalysts for our certain purposes. This, applied to the *de novo* design of artificial metallopeptides (Figure 5.9), will include:

- Selecting highly stable peptides that may be able to bear the inclusion of the non-canonical amino acid Ala-2,2'-bipyridine (BpyA) into their backbone and their structural characterization (Discovery stage, via Protein-ligand docking and MD simulations).

- Optimizing the identified stable complexes via rational re-design to enhance their catalytic activity and/or enantioselectivity for the conjugate addition of water to ketones (Optimization stage, MD simulations).

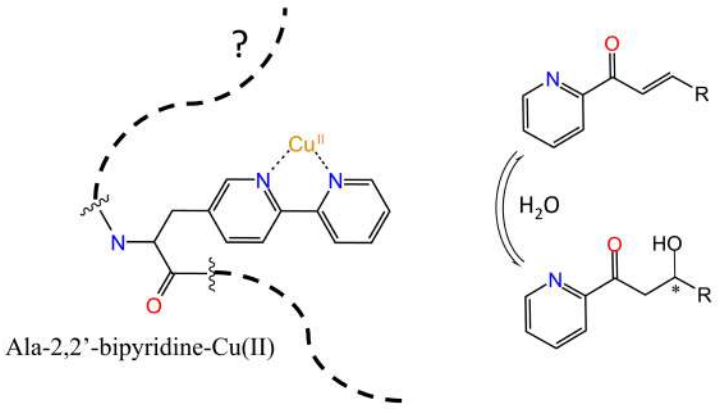


Figure 5.9: Initial stage for the *de novo* design of an artificial metallopeptide containing the UUA BpyA for the enantioselective hydration of ketones.

Successful designs will be synthesized by Vazquez's group at the University of Santiago de Compostela and catalytically tested by Roelfes and coworkers at the University of Groningen.

5.2.2 Computational details

Quantum calculations

The BpyA-Cu(II)-**2b** (substrate substituent R=t-butyl) complex was selected as model system (Figure 5.10), again in a bi-coordinated configuration. It was optimized with Gaussian 09 program [264] at DFT level using B3LYP-D3 functional. [200–202] The 6-31G(d,p) basis set was used [335–337] for non-metallic atoms and SDD [338] including a f polarization function for the 19 outer electrons of copper (the 10 inner core electrones are described by the SDD effective core potential (ECP)). The implicit Solvation Model based on Density (SMD) was used as continuum method. [286]

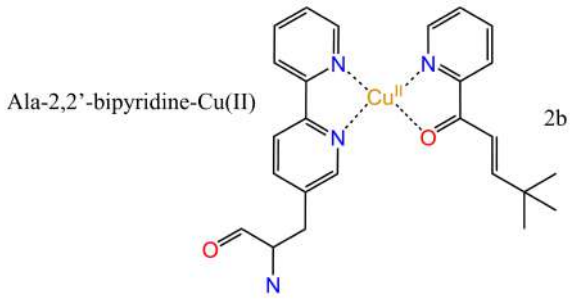


Figure 5.10: The copper bound non-canonical aminoacid Ala-2,2'-bipyridine (BpyA) coordinating the substrate **2b** (R=t-butyl) used in this work to model the artificial metallopeptides.

Protein-ligand docking

Protein-ligand docking was performed to introduce the optimized BpyA-Cu(II)-**2b** complex into different positions of each peptide generating a large variety of initial models. For that purpose, all these points were first mutated to A using the Dunbrack rotamer library [346] of the Chimera program [291] and then the BpyA-Cu(II)-**2b** complex was covalently linked at the selected points by overlapping the β carbon of the alanine with the corresponding carbon at the BpyA cofactor. The flexibility of the side chains of residues close to the linking point was considered to improve the adaptability of the peptidic scaffold to the BpyA-Cu(II)-**2b** complex.

Molecular Dynamics

The viability of the generated structures was assessed by submitting the best docking solutions to 100-500 ns MD simulation under the same conditions explained in previous section. In this case, the PBC were set up with a water box of 15 Å distance from the peptides to the edge of the box and the corresponding counterions (a variable number depending on the nature of each peptide) where used to neutralize the simulation cell. Those candidates that appeared stable were iteratively optimized and further submitted to new rounds of MD simulations up to 1-2 μs.

5.2.3 Results

***De novo* design: the discovery stage**

The first challenge to overcome at the initial stage of the designing process was to find suitable scaffolds to set up the starting point structures. Additionally, this was hampered by the specific requirements by the experimental group responsible for synthesizing the suggested peptides, which imposed initial restrictions on the nature of the designs. Initial structures should be:

- Highly stable to support the inclusion of the BpyA-Cu(II) cofactor without suffering denaturation.
- Rich on amino acids that could stabilize the BpyA-Cu(II)-**2b** complex, via either hydrophobic or polar interactions. Although stable peptides are then subjected to optimization, a big amount of changes in this stage should be avoided to guarantee the stability of the selected scaffolds.
- Small enough to achieve self-assembly (less than 30 amino acids).
- Disulphide bridges free, aiming to prevent crosslinking problems which could lead to unpredicted structures. Considering that S=S bonds are of the most powerful stabilizing motifs into proteins, this meant the main limitation of the designing process.

Based on these assumptions, a search of small and stable peptidic structures was performed into the Protein Data Bank (PDB). [28] The criterion for peptides selection was based on 1) their size, 2) their intrinsic nature regarding stability (β-sheets, cyclic backbone) and 3) the presence of specific amino acids that may be beneficial either for stabilizing the BpyA-Cu(II)-**2b** complex or for catalysis (D, E, W, Y or/and F). In fact, big and aromatic residues not only would provide stability but also could be hot spots to include the BpyA unnatural amino acid without promoting significant

distortion on the scaffold configuration. Taking all this into account, two different types of peptides were found as good candidates:

- The so called *WW domains*: considered the smallest natural motifs that are stable without containing disulfide bonds, they are small β-sheets found to be involved in natural protein-protein/protein-DNA recognition and protein signaling processes. [398,399] Interestingly, they have been successfully used for constructing new stable peptide variants [398,399] and for the study of protein-target recognition processes. [400] Additionally, they contain conserved W-W or W-Y hydrophobic regions which, combined with their stable geometry, makes them potential candidates to support the inclusion of the BpyA cofactor without damaging the structure.

- Small *cyclic peptides*: Alongside with the disulfide-crosslinked peptides, they are considered between the most stable structures to thermal and chemical denaturation. [24] The cyclic backbone imposes structural restraints that make the peptide to be less flexible, which could be beneficial to gain control on the stability of the structure after accommodating the cofactor. Additionally, these structural restraints could also facilitate managing the nature of the second coordination sphere at the optimization stage.

A set of six scaffolds was selected as starting point[2] (Figure 5.11): the WW domains 1zr7, [401] 2kbl, [402] 2m9e [403] and 2mwd; [404] the cyclized 2otq; [405] and the peptide 1e0n, [399] which is not related to the formers (neither WW domains nor cyclic) but share interesting structural features such as β-sheets and well-defined hydrophobic areas. Table 5.2 contains a brief description of the main structural properties of the selected peptides.

[2]The names of the peptides correspond with their PDB codes

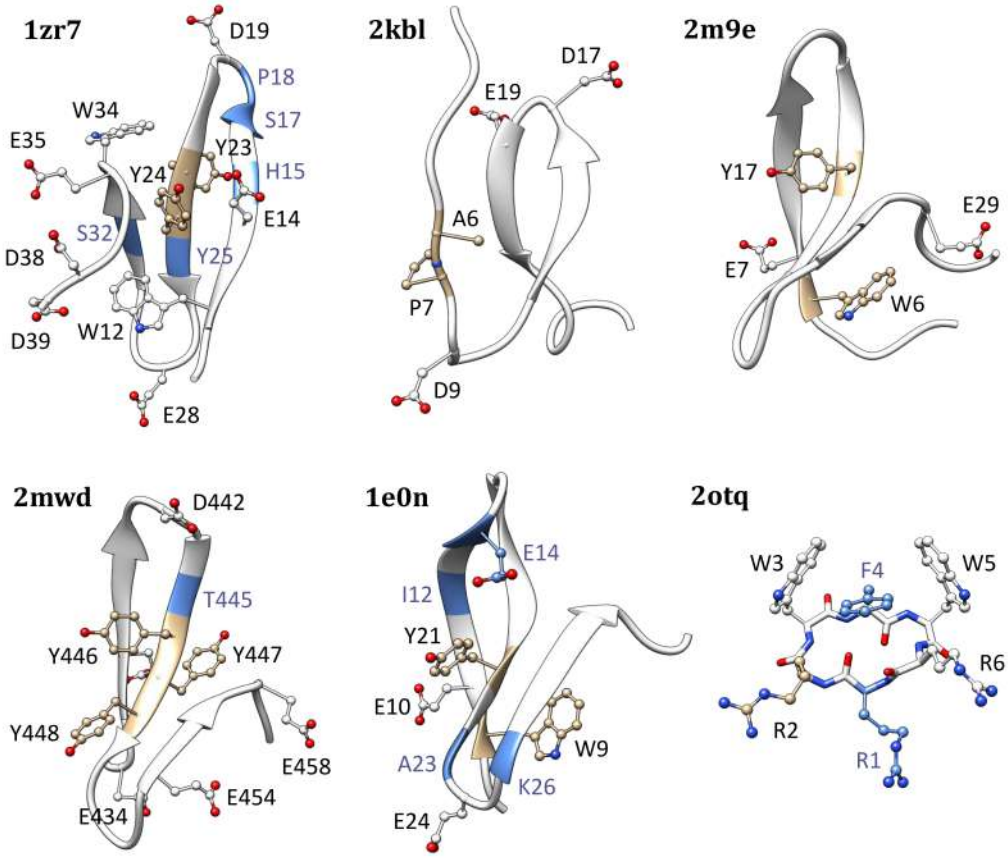


Figure 5.11: Illustration of the selected candidates from the PDB for the inclusion of the unnatural amino acid BpyA-Cu(II). Displayed residues correspond with the points where the BpyA-Cu(II) was included (brown), points with possible contribution to stability or catalysis (white) and points mutated along the optimization stage (blue).

Table 5.2: Structural properties of the peptides selected from the PDB. Consistently with Figure 5.11, labels correspond with mutated positions at the optimization stage (blue) and with points in which the BpyA-Cu(II)-substrate complex was included (brown).

Entry	PDB code	Type	Size	Sequence
1	**1zr7**	WW domain (Y23-W34) (Y24-W12)	30	GSWTEHKSPDGRTYYY NTETKQSTWEKPDD
2	**2kbl**	WWdomain (Y13-W20)	29	GSGYIPRAPRDGQAYV RKDGEWVLLSTFL
3	**2m9e**	WW domain (F11-F19-F28-Y17) (Y18-W6)	33	KLPPGWEKRMFANGTV YYFNHITNASQFERPSG
4	**2mwd**	WWdomain (W457-Y446-Y448) (Y447-W435)	27	SEWTERKTADGKTYYY NNRTLESTWEKP
5	**1e0n**	β-sheet	26	PGWEIIHENGRPLYYN AEQKTKLHYPP
6	**2otq**	cyclized	6	RRWFWR

After selecting the initial candidates, Ligand-protein docking was performed to include the optimized BpyA-Cu(II)-**2b** complex into a variety of positions of each peptide (Figure 5.11 and Table 5.2, labelled in brown). Thus, the six initial scaffolds leaded to the generation of twelve different variants (Table 5.3, see "Starting point" column).

From this stage on, the procedure varied depending on the specific nature of each system. In general, first, the stability of all the peptide-cofactor-substrate triads was assessed via 100-500 ns of MD simulations. After this time, those which underwent

denaturation were discarded and the stable ones were submitted to optimization as summarized in Figure 5.12.

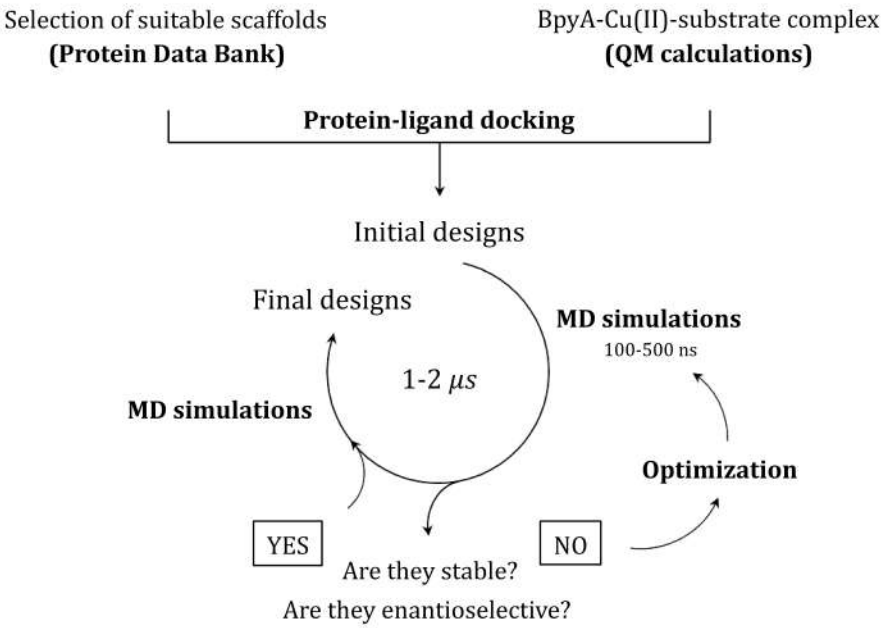


Figure 5.12: Description of the workflow implemented for the discovery and optimization of the BpyA-Cu(II) based artificial metallopeptides.

De novo design: the optimization stage

The estable peptide-BpyA-Cu(II)-**2b** candidates were subjected to optimization cycles aiming to confer them catalytic efficiency. These were based on:

- Generating stabilizing interactions with the BpyA-Cu(II)-**2b** complex. Either by promoting hydrophobic interactions with the BpyA-Cu(II)-**2b** complex (W or F) or placing coordinating residues close to the metal center to stabilize the orientation of the cofactor (E or G). Ensuring the orientation of the substrate together with an asymmetric environment results key for enantioselectivity.

- Enhancing catalysis by placing D or E residues as closer as possible to the substrate double bond.

Optimized structures were subjected to new rounds of MD simulation and, again, denatured variants were discarded. This iterative procedure of optimization plus MD simulation was elongated up to 1-2 μs until stable and catalytically promising variants were found. Table 5.3 includes all the tested peptides and the resulting candidates after the optimization stage.

Table 5.3: Peptide variants tested and selected for experimental validation.

Entry	PDB code	Starting point	Tested variants	Sel	Sequence	Final candidates
1	1zr7	W23BpyA	S32D	√	GSWTEHKSPDGRTXYY NTETKQDTWEKPDD	1A
			S32D_H15N	√	GSWTENKSPDGRTXYY NTETKQDTWEKPDD	1B
			S32D_H15W	X	GSWTEWKSPDGRTXYY NTETKQDTWEKPDD	-
			P18D	X	GSWTEHKSDDGRTXYY NTETKQSTWEKPDD	-
		Y24BpyA	-	X	GSWTEHKSPDGRTYXY NTETKQSTWEKPDD	-
2	2kbl	P7BpyA	-	X	GSGYIPRAXRDGQAYV RKDGEWVLLSTFL	-
		A6BpyA	-	X	GSGYIPRXPRDGQAYV RKDGEWVLLSTFL	-
3	2m9e	Y17BpyA	-	X	KLPPGWEKRMFANGTV XYFNHITNASQFERPSG	-
		W6BpyA	-	X	KLPPGXEKRMFANGTVYY FNHITNASQFERPSG	-
4	2mwd	Y446BpyA	S455D	X	SEWTERKTADGKTXYY NNRTLEDTWEKP	-
		Y447BpyA	-	√	SEWTERKTADGKTYXY NNRTLESTWEKP	4A
		Y448BpyA	-	√	SEWTERKTADGKTYXY NNRTLESTWEKP	4B
5	1e0n	Y21BpyA	I6W	√	PGWEIWHENGRPLYXN AEQKTKLHYPP	5A
			I6W_K20D	√	PGWEIWHENGRPLYXN AEQDTKLHYPP	5B
			I6W_K20D_E8N	X	PGWEIWHNNGRPLYXN AEQDTKLHYPP	-
			I6W_K20D_A17I	√	PGWEIWHENGRPLYXN IEQDTKLHYPP	5C
		W9BpyA	-	X	PGXEIIHENGRPLYYN AEQKTKLHYPP	-
6	2otq	R2BpyA	R1F_F4D	√	FXWDWR	6A

The selected candidates

The results point to variants **1A**, **1B**, **4A**, **4B**, **5B** and **6A** as the most promising candidates regarding stability and catalytic distances between the substrate double bond and a negatively charged residue. **5A** and **5C** variants show less stability than the formers and are expected to be less efficient. Still, all of them will be synthesized and tested experimentally to check the veracity of the computational models.

1zr7 based peptides **1A** and **1B** result the most promising to catalyse the hydration reaction (Figure 5.13). They show high stability during 1 μs MD simulation with no significant rearrangement from the original configuration after the inclusion of the BpyA-Cu(II)-**2b** complex (Figure 5.13, RMSD).

For both variants, the S32D residue stabilizes the cofactor by coordinating the metal center. At the same time, it maintains proper distances (around 5 Å) with the substrate double bond (Figure 5.13, distance graphs). In both cases, it is expected to boost the water nucleophilic activation to generate the *S* enantiomer. This configuration confers a double role to the aspartate residue: on one hand, increasing stability of the BpyA-Cu(II)-**2b** orientation and, on the other hand, boosting catalysis. In previous studies of this thesis, this arrangement has been related with systems with improved catalytic efficiency, regarding both conversion and enantioselectivity (See LmrR-Phen-Cu(II) in Chapter 4, and LmrR-BpyA-Cu(II)_V15E in the previous section of this Chapter). Additionally, the residue W12 is suggested to also contribute to the stabilization of the cofactor-substrate complex. It appears interacting via π-stacking with the aromatic ring of the substrate during all the MD simulation time scale.

The difference between the two variants is the position H15, which is mutated to N in **2b**. This mutation aims at avoiding possible competition with S32D to coordinate the metal center. The polar residue N was selected instead to replace the histidine with minimum electrostatic alteration.

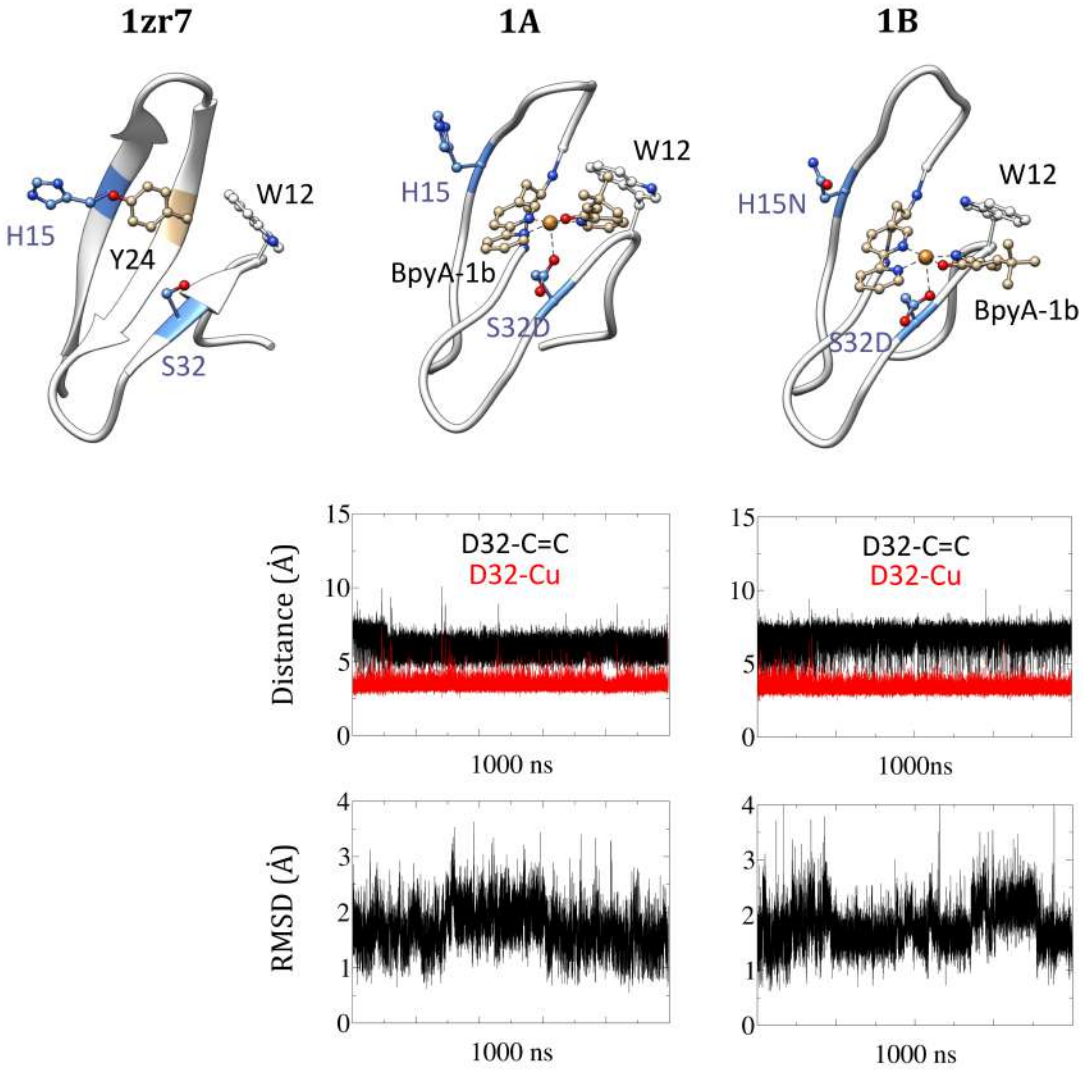


Figure 5.13: 1000 ns MD simulation results for the 1zr7 peptide variants 1A and 1B. In brown are labelled the positions in which the UAA BpyA-Cu(II)-substrate complex is included and in blue the residues that have been mutated during the optimization stage.

For the variants based on the 2mwd peptide **4A** and **4B** (Figure 5.14), the residue E454 is suggested to perform the water nucleophile activation. The distances between its oxygens and the substrate double bond are consistent with catalysis (about 5 Å). However, for both **4A** and **4B** the substrate does not appear packed enough, so it may acquire different orientations with respect to the copper catalyst. Thus, for these variants it is not possible to perform an estimation of the enantioselective tendency.

Still, hydrophobic interactions seem to play a crucial role stabilizing the BpyA-Cu(II)-substrate complex. For variant **4A**, W435 residue appears performing π-stacking with the bipyridine. However, this interaction seems not to be very stable and significant fluctuations of BpyA-Cu(II)-**2b** are observed (Figure 5.14, RMSD 4A). Regarding variant **4B**, the cofactor-substrate complex appears sandwiched between R438 and L453 residues, which maintain the complex better stabilized with respect to variant **4A** (Figure 5.14, RMSD 4B). Summarizing, some catalytic activity is expected for both variants **4A** and **4B**, with higher conversion levels related to **4b**. No clear enantioselective tendencies are observed.

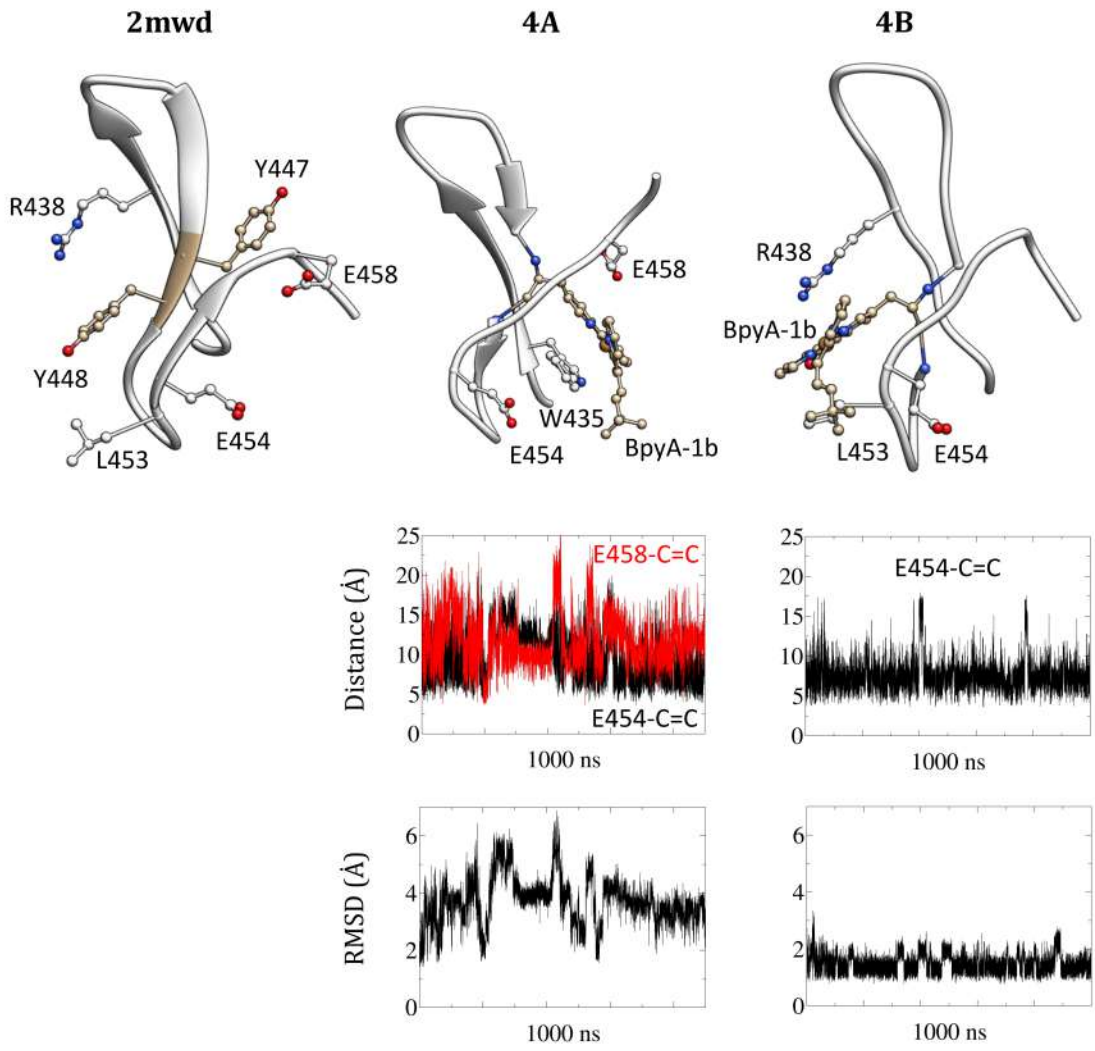


Figure 5.14: 1000 ns MD simulation results for the 2mwd peptide variants 4A and 4B. The positions in which the UAA BpyA-Cu(II)-substrate complex is included are labelled in brown.

For the 1e0n peptide, the variant **5B** results the most promising (Figure 5.15). K26D is found at catalytic distances (around 5 Å) with respect to the substrate double bond during 1.5 μs MD simulation (Figure 5.15, top). Even fluctuations in the global configuration of the peptide are found in the last 400 ns of simulation (Figure 5.15, RMSD), these interactions are still observed. Additionally, residue I12W seems to have an important role in stabilizing the cofactor-substrate complex next to K26D.

In this case, K26D is expected to approach the double bond for the generation of the S enantiomer.

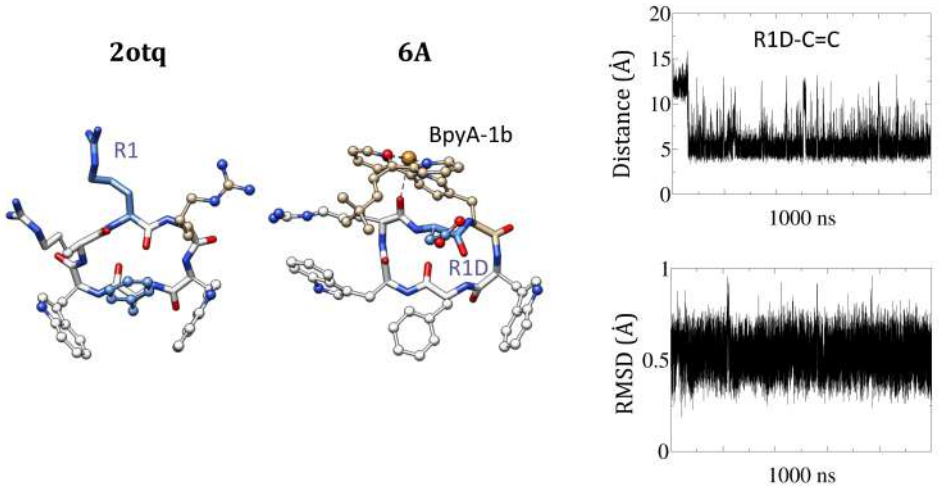


Figure 5.15: 1500 ns MD simulation results for the 1e0n peptide variant 5B. Displayed residues correspond with the position at which the UAA BpyA-Cu(II)-substrate complex is included (brown) and with points mutated during the optimization stage (blue).

Last, the 2otq variant **6A** is also between the most promising candidates (Figure 5.16). The BpyA-Cu(II)-**2b** complex change its placement at the very beginning of the MD simulation. In this new configuration, it becomes stabilized by the backbone carbonyl group of the residue R6, which coordinates the metal center. This arrangement, which is maintained along all the simulation time scale, leads to distances between the residue R1D and the substrate double bond of less than 5 Å (Figure 5.16, top). Regarding enantioselectivity, the substrate is not packed enough and opposite orientations with respect to the copper catalyst could be acquired. This would lead to a low or absent ee levels. Among the selected candidates, this peptide is expected to be the most efficient for the hydration reaction regarding conversion.

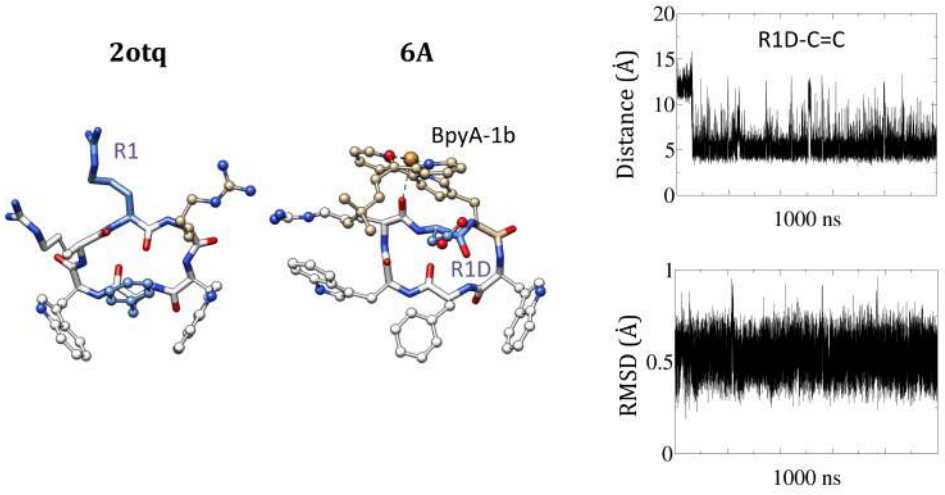


Figure 5.16: 1000 ns MD simulation results for the 2otq peptide variant **6A**. In brown is labelled the position in which the UAA BpyA-Cu(II)-substrate complex is included and in blue the residues that have been mutated during the optimization stage.

5.2.4 Conclusions and future perspectives

The modelling of artificial metallopeptides has been one of the most challenging procedures to overcome during this work. Starting from scratch, different peptides has been selected from the Protein Data Bank and re-designed to provide a proper environment for the unnatural aminoacid BpyA to perform the enantioselective hydration of ketones. The most promising peptides result those constituted by WW domains or with a cyclic backbone. A computational framework which combines Protein-ligand docking, Molecular Dynamics and iterative rounds of optimization has been successfully implemented to suggest a variety of peptides, from which at least four (**1A**, **1B**, **4B** and **6A**) are expected to be highly efficient. Hoping the designs are verified by upcoming experimental work (Vazquez's group for peptide synthesis and Roelfes's group for testing catalytic activity), these would be the first set of *in silico* designed artificial metallopeptides including a non-canonical amino acid able to perform the enantioselective conjugated addition of water to ketones.

Combining WW domains with metal chelating non-canonical amino acids appears as a promising approach for the design of new enantioselective systems and opens the door to its use for a different set of reactions mediated by copper for which enantioselective profiles are desired.

Chapter 6

Functional characterization of the active site dynamics in heme binding proteins

How protein dynamics modulate enzyme catalysis is currently one of the most active fields in enzymological research. A large variety of recent works relating the role of active site dynamics with the catalytic activity of proteins exists. [406–409] During the last years, there is a growing recognition that the X-ray structures do not provide with enough information to understand catalysis, which frequently requires global and/or local rearrangements of the scaffold. In many cases, elucidating this dynamical behaviour appears essential to understand catalytic properties such as selectivity and/or ligand specificity. [407, 410–412] Interestingly, many of these studies are driven by a computational input, in which MD simulations appear as a central element to assist the understanding of the interplay between protein conformation and function. [408–411, 413–415]

One of the weakness on enzyme design field lies precisely in underestimating (or the difficulty to estimate) the intrinsic dynamics of proteins, which nowadays is deemed

key for enzyme catalysis. [416–420] The poor consideration of protein dynamics in enzyme design is probably one of the bottleneck that makes the catalytic efficiency of new designs to be frequently many orders of magnitude lower than the one found in naturally occurring enzymes. [118–120]

The work comprised in next sections involves active site dynamics as key point to understand how either artificial or naturally occurring heme enzymes rearrange their structure to approach pre-catalytic states. First, an artificial heme enzyme, based again on the LmrR protein, developed by Roelfes and coworkers will be presented, as well as the multiscale strategy followed to elucidate the dynamical behaviour of the scaffold to approach pre-catalytic states. Second, two examples of naturally occurring heme enzymes, the latex cleavage protein (LCP) and the hemophore HasA, will be assessed to provide more light about the possible biochemical processes being key to drive active site dynamics and regulate both heme binding and/or catalysis.

6.1 An artificial heme enzyme for enantioselective cyclopropanation reactions

6.1.1 An overview

Engineered heme proteins have emerged as one of the most potential catalysts in the last years. [421, 422] Their versatility to catalyze new-to-Nature reactions including cyclopropanation, [423–427] olefination, [428, 429] N-H, [430, 431] Si-H [432] and B-H insertion [433] reactions, has converted them in principal targets for enzyme design. This includes from the optimization of naturally occurring enzymes [434–440] to the *de novo* design of artificial heme proteins based on 4-helix bundles or antibodies, between others. [87, 172, 392, 441–443]

As showed in previous Chapters, the LmrR protein owns a flat hydrophobic pore at the dimer interface that allows the accommodation of planar organometallic moieties. Roelfes and coworkers again used the LmrR protein as scaffold to incorporate the heme group via self-assembly. [171] The new artificial heme enzyme was able to catalyse the enantioselective cyclopropanation reaction (Figure 6.1).

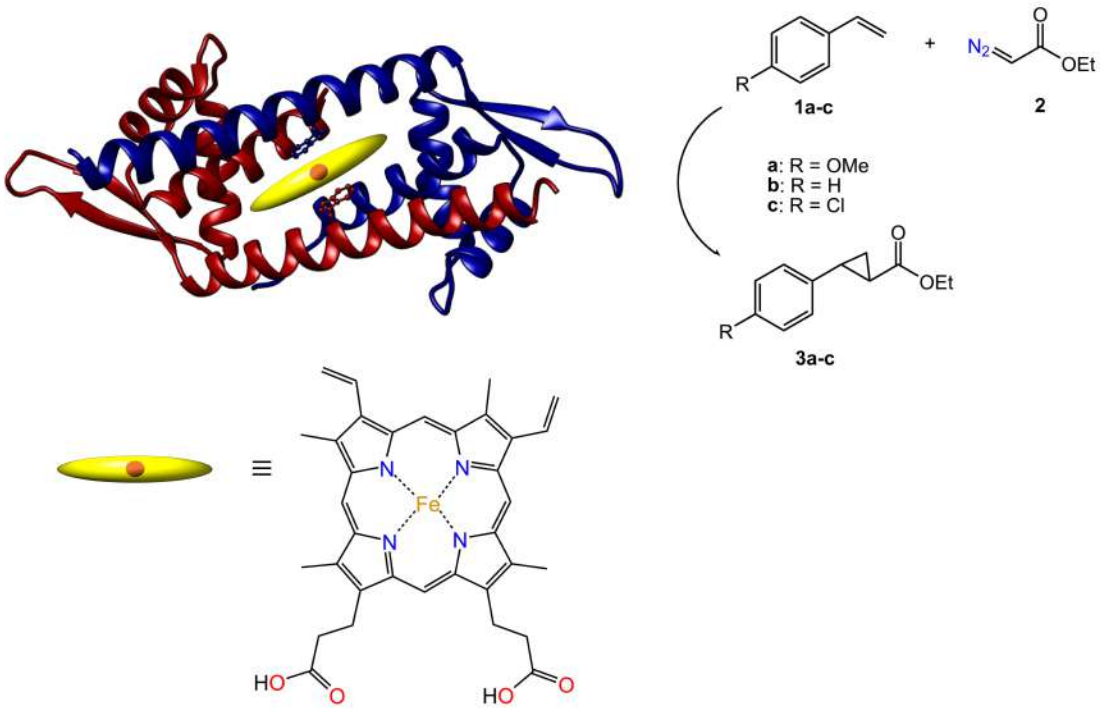


Figure 6.1: Illustration of the artificial heme enzyme designed by Roelfes and coworkers for the enantioselective cyclopropanation reaction. The heme group appears sandwiched between the residues W96/W96' at the center of the LmrR interdimeric region.

However, the solved 3D structure of the LmrR-heme enzyme (PDB code: 6UUF) shows a non-catalytic binding of the heme group at the dimer interface. In contrast to naturally occurring heme enzymes, which own a big hydrophobic cavity for accommodating the substrates, in the LmrR-heme system both axial faces of the porphyrin appear blocked by the indoles of both W96/W96' (Figure C.1). This configuration suggests no free space for the substrates to access the metal center and, thus, evi-

dences the necessity of the artificial system to experience structural rearrangements and acquire catalytic activity.

LmrR is known to be a quite flexible protein as this is key to carry out its biological function. [370, 444] This feature has been supported previously in this thesis, in which LmrR variants have experienced different rearrangements in response to mutations and to the inclusion of different cofactor-substrate partners. Based on this, we hypothesize that the dynamical behaviour of the LmrR-heme system could lead to "open" states that allow the incorporation of the different substrates.

A computational multiscale strategy which combines QM calculations, Protein-ligand docking and large scale MD simulations has been performed to study the ability of LmrR-heme artificial enzyme to accommodate different heme-substrate complexes. This encompasses, first, the identification of key structural features at the LmrR active site that allow the substrates to access to the metal center and, second, the analysis of the active site dynamics to identify the structural rearrangements that allow the accommodation of pre-catalytic structures at the dimer interface.

6.1.2 The experimental input

As a result of the cyclopropanation reaction of different styrene derivatives catalysed by the LmrR⊂heme system (Figure 6.1), the *trans* isomer *1R,2R* (trans/cis = 92:8) was obtained as major product with a 25% yield and an ee of 17% (Table C.1).

To further asses the role of the protein scaffold and trigger optimization, a variety of mutations were performed in the residues located at and around the active site. Residues W96, F93, V15, M8 and D100 were mutated to A and further tested. Excepting V15A and W96A, all variants showed an enhanced catalytic activity, comparable or greater than the one provided by the LmrR itself. For LmrR_W96A, a complete loss of ee was found. This was attributed to a LmrR_W96A-hemin interaction that could be away from the hydrophobic cavity, resulting in catalytically active

assemblies but without defined chiral interactions to induce enantioselectivity. For mutants D100A and M8A, a notable increase in activity was found with the latter also providing the highest ee (51%).

Due to the impossibility to rationalize the experimental observations based on the X-ray structure, which shows a non-catalytic assembly of the heme into the protein cavity, computational studies were performed to gain more insight about the origin of the catalytic activity found in the LmrR⊂heme system.

6.1.3 Computational details

Quantum calculations

The heme complexes used to study the dynamical behaviour of the LmrR⊂heme ArM were: 1) heme (**5**), 2) heme-carbene (**6**) and 3) the transition state of the cyclopropanation reaction to approach the *1R,2R* product (**7**) (Figure 6.2). These were obtained via QM calculations with the Gaussian 09 program [264] at DFT level using B3LYP-D3 functional. [200–202] The 6-31g(d,p) basis set [335–337] was used for non-metallic atoms and the SDD [338] including a f polarization function was used for the iron. The singlet state was considered for the metal center.

The implicit Solvation Model based on Density (SMD) was used, [286] considering an epsilon adequate to describe solvent accessible surfaces of proteins (ε=9). [339]

Protein-ligand docking

The crystal structure of LmrR protein bound to the heme group (PDB code: 6FUU) was used as model system to dock the different complexes **5**, **6** and **7**. Crystallographic water molecules were removed from the model. Side chain flexibility was considered for those residues pointing towards the hydrophobic cavity using the Dunbrack rotamer library. [346, 358] All docking runs were performed with GOLD 5.2 (available

through the Cambridge Crystallographic Data Center (CCDC)), and evaluated with ChemScore [256] scoring function.

Molecular Dynamics

The best scored docking solutions were used as starting point models for all-atoms molecular dynamics (MD) simulations. The loops composed by the residues 70-73 in both chains A and B, not determined in the X-ray experiment, were constructed by superimposition with previous models. Terminal residues 1-4 and 108 of chains A and B, that were not solved neither, were discarded and uncharged terminal motifs NME and ACE (as implemented in AMBER force field) [208] were used to end the chain terminals. H86, solvent exposed, was considered protonated at the ε position. Models were set up with the xleap program. [208] Each system was embedded into a cubic box including about 37000 water molecules and 4 sodium ions to neutralize the simulation cell. The AMBER [208] and TIP3P [347] force fields were used for protein and water, respectively. For sodium ions, parameters from ions94.lib library were used.

Parameters for the LmrR bound to **5**, **6** and **7** were developed. Point charges were calculated with antechamber [445] based on the RESP procedure. [305] Bonded terms at the Fe center were calculated according to Seminario's method based on second-derivatives. [302] The GAFF force field [223] was adopted for the remaining atoms.

A cutoff of 10 Å was used for short range electrostatics and Van der Waals interactions. Long range electrostatic interactions were calculated with the Particle Mesh Ewald method. [283] Bonds involving hydrogen atoms were constrained using the SHAKE algorithm. [348] A time step of 1 fs was used to integrate the equation of motion with a Langevin integrator. [349, 350] The standard protocol followed to perform the MD simulation was the same as the used in previous works of the the-

sis: constant temperature and pressure were achieved by coupling the systems to a Monte Carlo barostat at 1.01325 bar. [351] Model systems were initially energy minimized (3000 steps) progressively, allowing water molecules, side-chain and backbone atoms to move; then, thermalization of water molecules and side chains was achieved by increasing the temperature from 100 K up to 300 K; finally, 100 ns MD simulations under NPT conditions were performed and further analysed. Simulations were performed with OpenMM 7.0. [352] and molecular visualization and graphics were performed with the UCSF Chimera program. [291]

6.1.4 Results

We first performed a purely QM study of the cyclopropanation reaction in the absence of the protein to get distinct intermediates (the heme alone, labelled as **5** and its adduct with the carbene intermediate, labelled as **6**) and the TS structure that leads to the generation of the *1R,2R* cyclopropane (the predominant product found experimentally, labelled as **7**) (Figure 6.2 and 6.3).

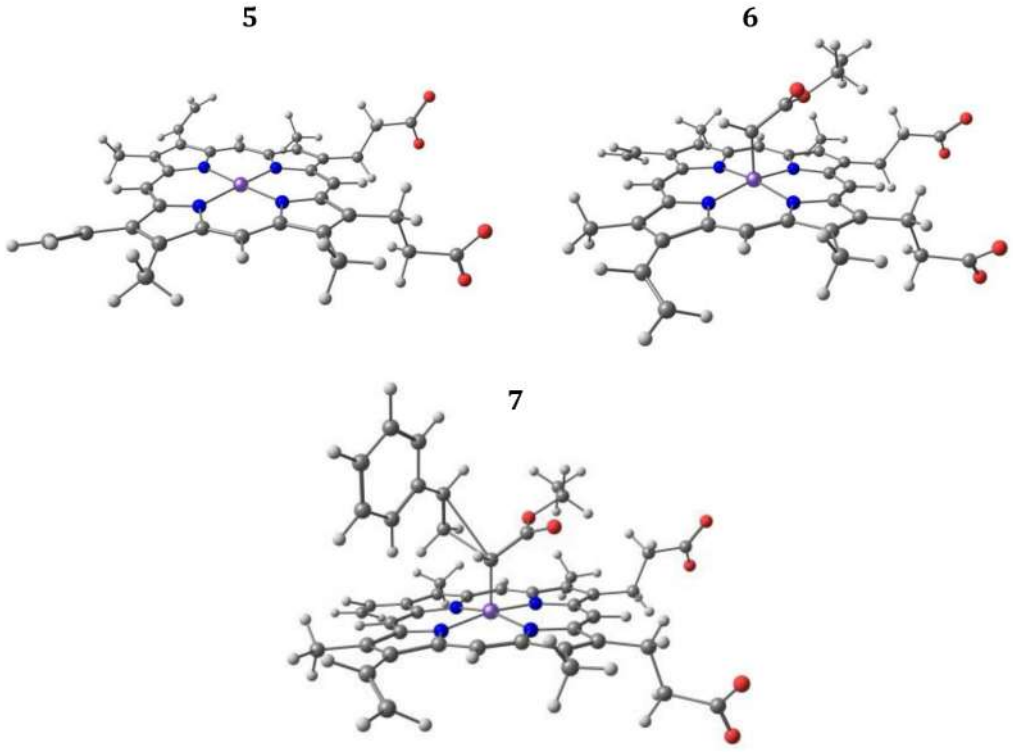


Figure 6.2: Structures resulting from QM calculations and used as ligand for Protein-Ligand Docking. The heme group alone (**5**), its adduct bound to the carbene substrate (**6**) and the TS structure (**7**).

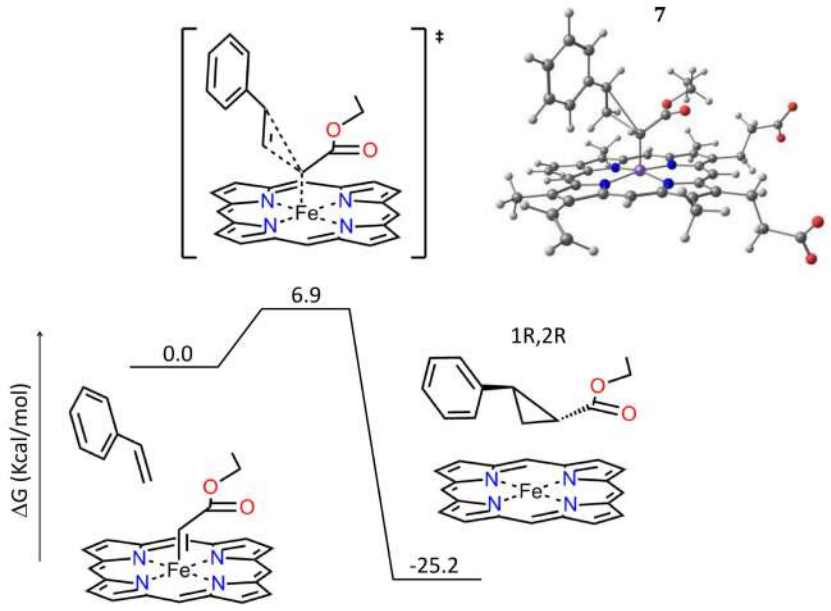


Figure 6.3: Energy profile and TS geometry (**7**) for the last step of the cyclopropanation reaction.

Protein-ligand docking of both (**5**) and (**6**) moieties were carried out on LmrR, starting from the LmrR⊂heme X-ray structure with the heme group removed. For both **5** and **6** four different binding poses with good predicted affinity (around 50 ChemScore units) were found (Table C.2), corresponding to rotations of the heme group around the orthogonal axis with respect to the average plane of the heme passing through the metal. The best docking solutions of LmrR⊂**5** showed the heme sandwiched between W96/W96', consistently with the crystal structure (Figure C.2 a, b). Since the residue W96/W96' appears key for the enantioselective capacity of the enzyme, the heme group was also docked into the variant LmrR_W96A to assess how this mutation could affect the binding of the heme group. In contrast to the WT LmrR, in which all the docking poses resulted superimposed, this variant provided a high variety of distinct orientations of the heme group at the hydrophobic cavity (Figure C.2 c, d), supporting the experimental observation that the W96/W96' residues are essential for the packing of the heme moiety into the LmrR. Regarding the LmrR bound to the heme-carbene complex (LmrR⊂**6**), low energy solutions showed the W96' residue flipped towards the solvent to provide free space for the accommodation of the co-substrate at the active site (Figure C.2 g, h).

To assess the evolution over time of the different systems and elucidate possible structural rearrangements, the four predicted structures of LmrR⊂**5** and LmrR⊂**6** (a total of eight structures) were submitted to 100 ns MD simulations.

Cluster analysis of the merged trajectories for both LmrR⊂**5** and LmrR⊂**6** systems shows a high stability of the heme moiety at the active site, with W96/W96' residues generating a hydrophobic patch that packs the heme. To better analyse the entire MD trajectory in term of individual frames, a cluster analysis was performed (Figure C.3). Between the most populated clusters of the LmrR⊂**5** system, some (clusters 1 and 3-6) show high similarity with the X-ray structure, with the metal center inaccessible to the solvent and, by extension, to the substrates (Figure 6.4 a,

and C.4 a). Other clusters though (0 and 2), show changes in the conformation of the helix containing the W96' residue (α4) and the indole of W96' reoriented towards the entrance of the cavity. This is consistent with previous reports which state that flexibility in the helix α4 is a major contributor for the LmrR to adapt its structure to different drugs. [444] These changes are accompanied by a significant displacement of the heme moiety towards the entrance of the active site, which results in an "opened" conformation with significant free space at one axial face of the porphyrin. This appears to be key for the metal center to be accessible to the substrates and the catalysis to proceed (Figure 6.4 b, and C.4 b).

The MD simulations of the LmrR bound to the heme-carbene (LmrR$\subset$**6**), support the role of the W96' residue for the complex to accommodate at the active site: when W96' appears pointing towards the hydrophobic core, the heme-carbene complex is directed towards the solvent since there is not free space for the carbene to access the active site (Figure 6.4 c, and C.4 c). In contrast, those trajectories resulting from starting points in which the docking accommodates the ligand by rotating the W96' residue, the heme-carbene complex (**6**) remains properly located at the dimer interface (Figure 6.4 d, and C.4 d).

Last, we analysed the effect of the second substrate (styrene) on the conformation of LmrR scaffold. For that purpose, we use as model system the TS geometry of the cyclopropanation reaction (complex **7**), as this is the most bulky intermediate found along the catalytic pathway.(Figure 6.3) Protein-ligand docking was performed into the LmrR scaffold then subjected to the same analysis as before.

Docking results show good affinities (about 48 ChemScore units) between the complex and the active site residues (Table C.2). However, in this case, all the structures present the W96' pointing towards the cavity. This makes the heme-TS complex to be displaced towards the entrance of the cavity to accommodate the bulky TS (Figure 6.5, left). The best scored pose was submitted to 100 ns MD simulation

and cluster analysis. This revealed an alternative rearrangement of the LmrR scaffold: the majority of the structures found (clusters 0, 2 and 3) suggest a broader dimer interface in which the two helix containing the W96/W96' residues appear further separated to accommodate the catalytic complex at the hydrophobic cavity (Figure 6.4 e, and C.4 e). In this state, W96' appears to be key in stabilizing the TS geometry via π-stacking with the phenyl group of the styrene.

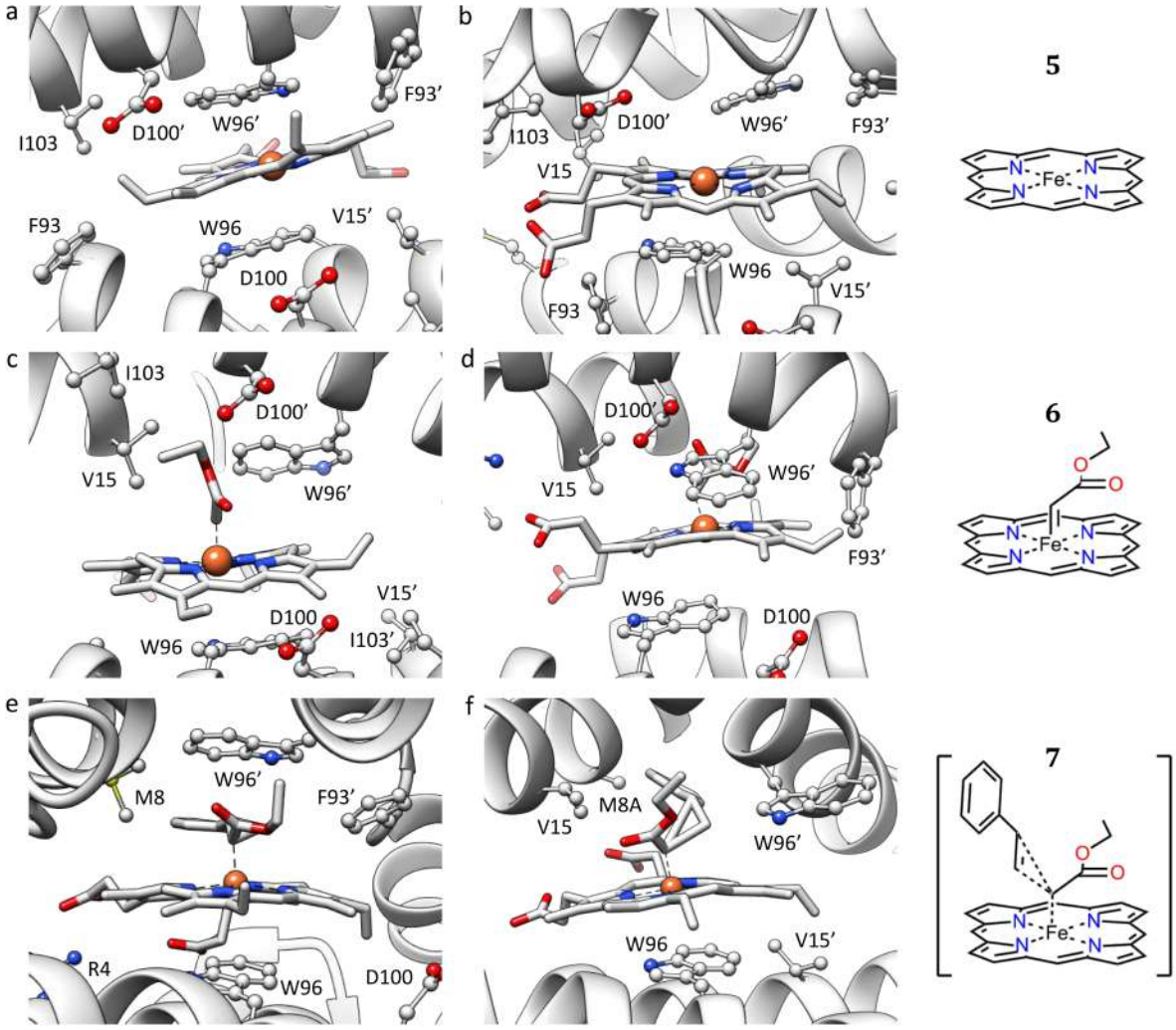


Figure 6.4: Representative structures resulting from MD simulations of: the LmrR⊂heme (**5**) system (a) cluster 3, RMSD with crystal of 1.634 Å and b) cluster 2, RMSD with crystal of 2.471 Å. 400 ns MD simulation), the LmrR⊂heme-carbene (**6**) system (c) cluster 0, RMSD with crystal of 0.548 Å and d) cluster 2, RMSD with crystal of 1.000 Å. 400 ns MD simulation) and the heme bound to the cyclopropanation transition state (**7**) (e) into LmrR, f) into LmrR_M8A. 100 ns MD simulation).

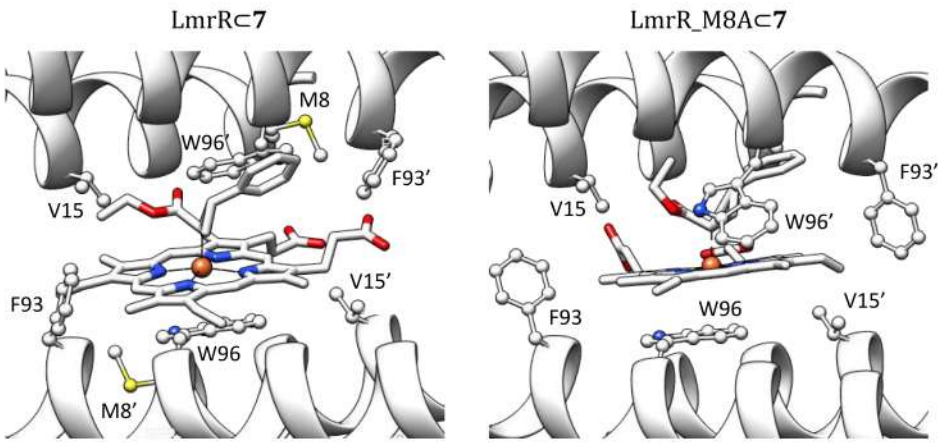


Figure 6.5: Best docking solutions of the heme-TS complex (**7**) into the LmrR scaffold (left) and its variant LmrR_M8A (right).

Since the M8A mutant showed the best results in catalysis, the dynamic behaviour of this variant was also assessed. The system resulting from the docking of the heme-TS (**7**) complex into LmrR showed a clear difference with respect to the WT LmrR: in this case all the poses resulted with a W96' pointing towards the solvent. This suggests that the reduced size of the M8A residue allows allocating the heme-TS (**7**) complex into the active site (Figure 6.5, right). The best docking solution was selected as starting point for a 100 ns MD simulation. The results suggest that the effect of the M8A mutation is mainly steric: the aromatic moiety of the styrene now occupies the free space resulting from both the change of the bulky methionine into the much smaller alanine and the rotation of the W96' towards the solvent, enhancing again the role of this residue to approach conformations compatible with catalysis (Figure 6.4 f).

6.1.5 Conclusions and final remarks

The combination of experimental work and computation demonstrates that the cyclopropanation reaction occurs unequivocally at the hydrophobic cavity of the LmrR protein. Initially, this was counterintuitive since X-ray experiments showed a LmrR

pocket fully occupied upon binding of the hemin, whereas heme enzymes (such as P450) usually present a large hydrophobic cavity for binding substrates. Yet, the artificial metalloenzyme exhibited good activity and enantioselectivity. Computation shows that this ability is attributed to the dynamic nature of the LmrR⊂heme complex, which involves significant geometric rearrangements of the heme environment allowing binding of the substrates. The residue W96/W96' located at helix $\alpha 4/\alpha 4$' was identified as the main responsible to allow the heme-substrate complexes to reach the center of the hydrophobic cavity. The combination of QM calculations, Protein-ligand docking and MD simulations leaded to the identification of transient open conformations which allow allocating the substrates at the active site and reaching pre-catalytic states for the reaction to proceed.

6.2 Structure-function relationship in naturally occurring heme binding proteins

Metal mediated recognition processes are of the most complex questions that molecular biology has to address. The biological-inorganic interplay overcomes standard knowledge in both chemical and biological sciences and the assessment of the relative contributions of both partners in the recognition mechanism is extremely challenging. [446]

The heme is the most ubiquitous naturally occurring inorganic ligand and, therefore, represents a prototypical system to decode the organometallic binding to proteins. Due to the versatility of heme mediated catalysis, in the last years ArM design has stand on heme binding recognition to approach new-to-Nature reactions. [57, 171, 172, 447] However, the common knowledge about the heme binding process to hosts, regarding active site dynamics, the interplay between the protein-cofactor-

substrate triad and the sensitivity of all these phenomena to environmental magnitudes is still quite limited.

Recent studies on heme binding proteins have related changes in the media and/or certain residues of the protein with structural rearrangements that can affect the substrate recognition process and the catalytic activity. [413, 448, 449] Specifically, modifications on the protonation state of amino acids at the active site can promote structural rearragements key for the regulation of heme mediated processes. [450] In the last years, several studies show the potential of MD simulations to access to these structural rearrangements and identify active vs inactive conformations at the heme binding sites. [413, 448, 449, 451–453]

Since heme is so frequently found in Nature and it is increasingly used in the enzyme design field, we decided to move forward in this PhD by working on simulations on naturally occurring heme binding proteins that could shed light on bioinorganic recognition processes. In this study we combine Protein-ligand docking and large-scale MD simulations to study the interplay between active site dynamics and the regulation of heme recognition and heme based catalysis. For that purpose, we decided to focus on systems which involve the coordination of the metal center by the protein. These were: 1) The hemophore HasA, to asses, on one hand, how the heme binding can be mediated by the structural accessibility of the *apo* cavity and, on the other hand, how this process can be dependent on changes in the protonation state of the active site amino acids. 2) The Latex Cleaving Protein (LCP), a system that exemplifies very well the interplay between receptor dynamics, heme binding and substrate accessibility. We consider this is a good candidate to gain more insights about the strategies for modulating both heme binding and heme mediated catalysis at the protein binding site.

6.2.1 Computational Details

Model systems

The X-ray structures of HasA (PDB code: 1dkh) and both the *apo* and *holo* forms of LCP (PDB codes: 5o1m and 5o1l, respectively) of the heme-binding proteins were taken from the Protein Data Bank (PDB). [28] The crystallographic waters were removed from the model. Duplicated conformers of amino acids were replaced by consistent rotamers from the Dunbrack rotamer library [346] of Chimera program. [291]

Protein-ligand docking

Docking calculations of the heme moiety into the X-ray HasA and LCP proteins were carried out with GOLD5.2 program. [236] The side chains flexibility was taken into account using the rotamer library implemented in the same program. An adapted version of the GoldScore scoring function able to consider interactions between the protein and the first coordination sphere of the metal was used for evaluation of the poses. [454] The solutions were analysed with GaudiView. [455]

Molecular Dynamics simulations

Molecular dynamics (MD) simulations were set up with the xleap program. [208] The proteins were solvated with cubic box of water molecules using the TIP3P force-field. [347] Sodium ions were added to neutralize the simulation cell using the ions94.lib library. The force-field terms involving the first coordination sphere of the metal center were calculated via standard approaches. The force constants and equilibrium parameters were obtained through the Seminario's method [302] from a minimum considering the singlet state for the metal. Calculations were performed with the Gaussian 09 program [264] at DFT level using B3LYP-D3 functional. [200–202] The

6-31g(d,p) basis set [335–337] was used for non-metallic atoms and the SDD [338] including a f polarization function was used for the iron. The rest of parameters involving non-metallic atoms were obtained from the GAFF force-field. [223] The point charges were calculated with antechamber [311] using the RESP model. [305]

The cutoff for the Van der Waals and short-range electrostatics interactions was set to 10 Å. The long-range electrostatic interactions were calculated via the Ewald Particle Mesh method. [283] The SHAKE algorithm [348] was used to constrain bonds engaging hydrogen atoms. The equation of motion was integrated with a Langevin integrator [349, 350] and using a time step of 1 fs. To get constant temperature and pressure the system was coupled to a Monte Carlo barostat at 1.01325 bar. [351]

OpenMM 7.0 program [352] was used to run all the MD simulations. A 3000 steps energy minimization was initially performed to progressively allow accommodating the water molecules, side chains and the backbone; then, the temperature was increased from 100 to 330 K to allow thermalization of water molecules and side chains; after that, 100 ns of production were carried out and further analysed.

6.2.2 The case of study of the latex cleaving protein LCP

An overview

The latex clearing protein LCP has a crucial role in cropping latex molecules and is known to regulate the process by promoting conformational changes at the active site, which involve the reorientation of coordinating residues. Solved 3D structures show that the main protagonist in this regulation is the K167 residue, which is placed at one face of the porphyrin plane (Figure 6.6) and is able to control the access of latex molecules to the metal center throughout its direct coordination. According to X-ray structures, this interaction with K167 is accompanied by a closed arrangement of the loop successive to this helix, which appears blocking the entrance of the heme bound cavity. This is known as a “closed” (inactive) conformation (PDB code: 5oim)

(Figure 6.6 a). In contrast, K167 is also able to acquire an "opened" (active) state (PDB code: 5oil), in which it is oriented towards the solvent and makes the metal center accessible to latex molecules. In this case, the loop does not appear blocking the entrance of the pore anymore, but is reoriented acquiring an opened state (Figure 6.6 b).

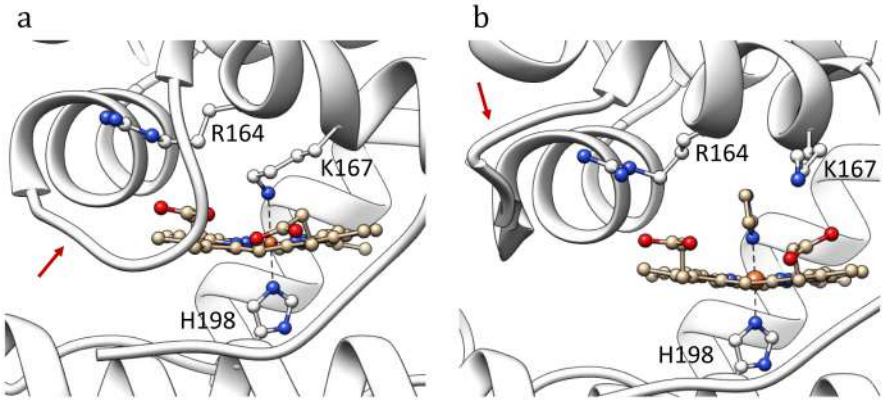


Figure 6.6: X-ray structures of the a) "closed" (PDB code: 5oim) and b) "opened" (PDB code: 5oil) states of the LCP protein. The red arrow is pointing to the loop which mediates the opening or closing of the entrance of the cavity.

Results and discussion

100 ns MD simulation of the *apo* form of LCP protein in its open state suggests that this state is highly stable in absence of the heme group. During the simulation time scale no transition from the opened to the closed state is observed (Figure 6.7 a, b). However, it is important to remark that the helix containing the lysine K167 shows significant flexibility during the simulation in contrast to the rest of the system, appearing highly distorted with respect to the X-ray (Figure 6.7 b). Clustering analysis supports this observation showing several conformations related to movements of the helix, which suggests a possible role in the transition from the closed to the opened state or vice versa.

Aiming to check if the protonation state of the lysine K167 could have any role in promoting these conformational changes, the same starting point but with a de-

protonated lysine was considered. After 100 ns of MD simulation, similar dynamic tendencies between the two systems were found. In this case, MD simulations were not able to relate the protonation state of the K167 with the modulation of the transition from opened to closed states.

With respect to the 5o1l (closed) system, 100 ns MD simulation of the *apo* form were also carried out. Although significant rearrangements concerning the backbone of the protein are not observed, substantial changes are found at the final stage of the trajectory regarding the position of the lysine K167, which appears reoriented towards the solvent. Thus, in this case the MD simulation was elongated during 50 ns. After this time, rearrangements in the backbone were elucidated. The loop that initially appeared blocking the entrance of the pore transits from the closed to the opened conformation (Figure 6.7 c). Representative structures of the acquired open state, significantly differ from the closed 3D structure (5o1m) used as starting point (Figure 6.7 d). In contrast, the new state appears quite similar to the 3D structure related to the opened state (5o1l) (Figure 6.7 e). All this together suggests that, in absence of the heme group, the protein tends to transit from the closed to the opened state.

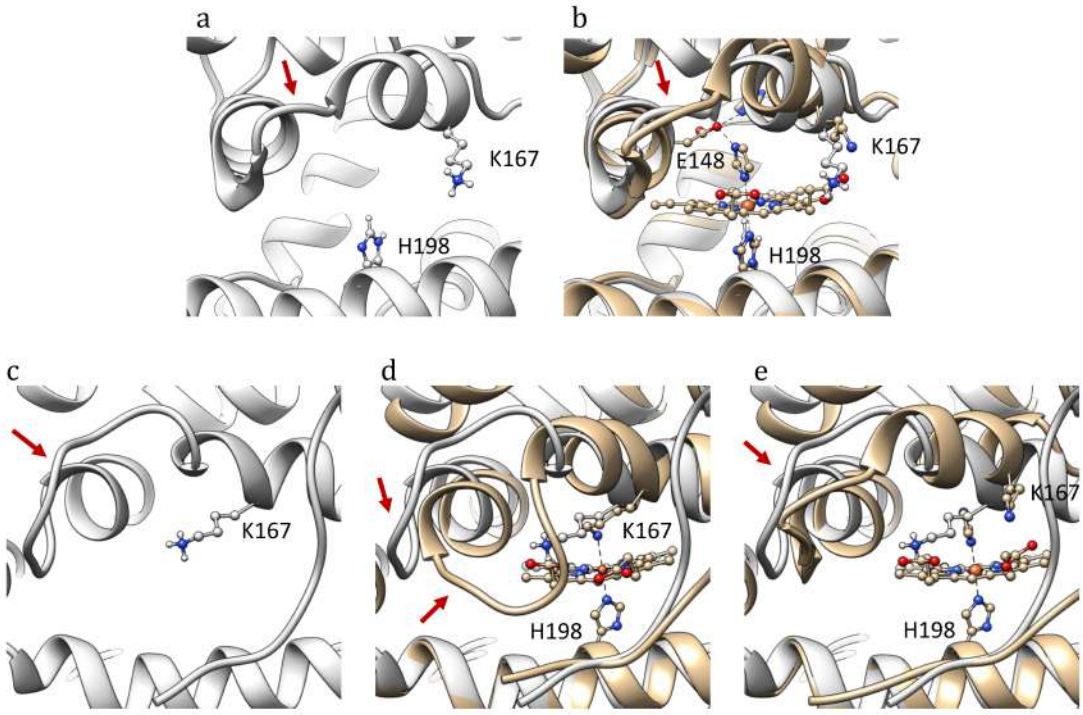


Figure 6.7: Representative structures of a) 100 ns MD simulation of the *apo* protein (in grey), starting from the opened state (5o1l), both superimposed in b) being the X-ray structure represented in tan; and c) repressentative structure of 150 ns MD simulation (in grey), starting from the closed state (5o1m), superimposed with the X-ray (in tan) corresponding to d) the closed state (5o1m) and e) the opened state (5o1l).

To further assess the possible role of the protonation state of K167 residue in this transition, the 3D structure corresponding to the closed arrangement (5o1m) was used as starting point, but in this case with lysine K167 in a deprotonated state. Surprisingly, after 100 ns of MD simulation, the backbone maintained the closed conformation (Figure 6.8). To ensure about the stability of this state, the MD simulation was elongated during 100 ns more, time after which the protein continued in a closed arrangement. Further analysis about the nature of the interactions occurring at the active site showed that the deprotonated K167 is able to perform transient interactions via hydrogen bonds with H167 (Figure 6.8 a, and C.5) and R133 (Figure 6.8 b, and C.5), which seems to contribute to maintain the lysine oriented towards the

cavity. It seems that this disposition of K167 prevents cascade effects occurring at the helix containing this lysine, and results in the maintenance of the closed conformation of the loop.

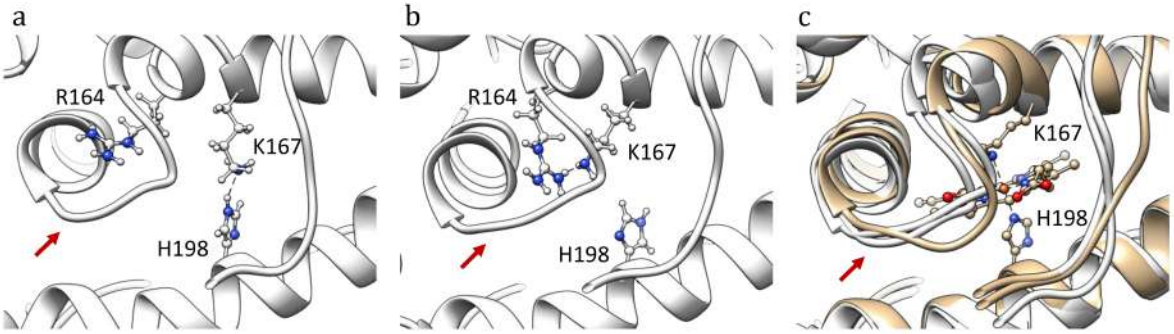


Figure 6.8: 200 ns MD simulation of the *apo* form of 5o1m (closed) with a deprotonated K167 residue. Stabilizing interactions are found along the simulation time scale both with a) H198 and b) R164. In c) the frames (in grey) are superimposed to the X-ray structure (in tan) used as starting point.

Conclusions

Results suggest that both lysine K167 and the loop located at the entrance of the active site could act in synergy to block the catalytic activity of LCP: when the deprotonated K167 chelates the metal center, first, this makes it inaccessible to latex molecules and, second, this avoids cascade effects preventing the loop to acquire a opened state.

These conformational rearrangements could also have an important role in the binding of the heme group: MD simulations of the closed *apo* forms containing either a deprotonated or a protonated K167, suggest an opposite behaviour regarding the opening or closing of the entrance of the pore. While with the protonated K167 the *apo* cavity tends to acquire an opened arrangement, the deprotonated K167 remains oriented towards the cavity preventing the opening of the loop and, by extension, the accessibility of the porphyrin. This suggests that changes on environmental factors

that may affect the protonation state of K167, could be key for the accessibility of the heme group to the active site of the LCP protein.

These results suggest that the protonation state of the amino acids located at the binding sites of natural heme binding proteins can play an important role both in the binding of the porphyrin as well as in the regulation of catalysis. Also, that this regulation can be governed by conformational changes of the heme binding cavity. Again, the combination of Protein-ligand docking and MD simulations has demonstrated to be a good approximation to assess the rearrangements resulting from slight changes occurring at the active site, in this case to characterize the heme binding into the LCP protein.

6.2.3 The case of study of the hemophore HasA

An overview

HasA is the first hemophore that was described at molecular level. [456] It is secreted by the Gram negative bacteria *S. marcercens* to the extra-cellular media under iron deficiency conditions to extract the heme group from hemoglobin. [457] Then, a synergistic incorporation of heme group into the bacteria is performed by increasing the efficiency of the membrane receptor HasR, which releases the heme into the bacterium to be used as iron source. The coordination of the heme group at the active site of the hemophore is associated with low redox potentials. Both this and the stability of the heme at the active site have been demonstrated to be enhanced when the coordinating histidine is partially or totally deprotonated. [457] These features makes the HasA hemophore a good model system for our purposes based on assessing if the chemical properties of the active site can affect its dynamical behaviour and, therefore, the binding of the heme group.

The X-ray hemophore HasA (PDB code: 1dkh) was used as initial model system. It contains two large loops able to pack the heme group and coordinate the metal

center through interactions with histidine H32 at loop A and tyrosine Y75 at loop B. Also hydrophobic interactions (V37, F45, L50, L77 and F78) and hydrogen bonds between N34 and one of the carboxylic groups of the porphyrin seems to stabilize the binding (Figure 6.9).

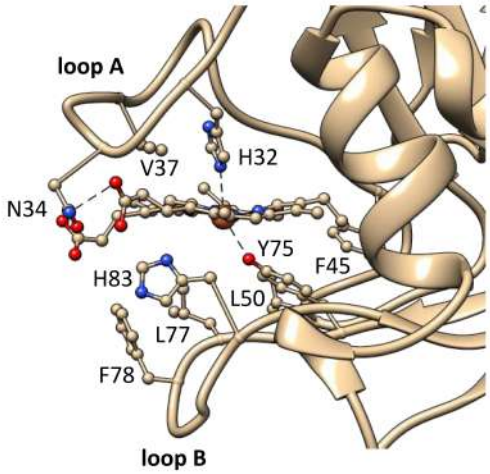


Figure 6.9: X-ray structure of the hemophore HasA (PDB code: 1dkh).

Aiming to describe the dynamical behaviour of the active site and to assess the possible influence of the protonation state of the coordinating amino acids (H32 and Y75) in the structural arrangement of the active site, the *apo* form of the HasA protein was considered in different protonation states: 1) histidine protonated in epsilon H32$^{\varepsilon}$ and tyrosine protonated Y75^{H} (system H32$^{\varepsilon}$-Y75^{H}), 2) histidine protonated in delta H32$^{\delta}$ and tyrosine protonated Y75^{H} (system H32$^{\delta}$-Y75^{H}) and 3) histidine protonated in delta H32$^{\delta}$ and tyrosine deprotonated Y75 (system H32$^{\delta}$-Y75).

Results and discussion

For system 1) H32$^{\varepsilon}$-Y75^{H}, the *apo* form of HasA was not stabilized until 40 ns of simulation due to significant rearrangements of the loops A and B. In absence of the porphyrin, these loops interact together acquiring a closed arrangement. Residues H32$^{\varepsilon}$ and tyrosine Y75^{H} appear interacting through a hydrogen bond (Figure 6.10 a,

b). This conformation is additionally stabilized by both a hydrogen bond between the tyrosine $Y75^{H}$ and the carbonyl group of the backbone of the glycine G33, and hydrophobic interactions between the amino acids that were previously stabilizing the heme group. Cluster analysis reveals that this geometry is significantly the most representative during the last 60 ns of MD simulation (Figure C.6).

For system 2) $H32^{\delta}$-$Y75^{H}$, the change in the protonation state of H32 does not allow its interaction with $Y75^{H}$, so this histidine appears flipped towards the solvent (Figure 6.10 c, d). Instead, $Y75^{H}$ interacts with the carbonyl group of the backbone of the G33, in loop A. This interaction seems to help to maintain the loops in a closed arrangement even lost the hydrogen bond between $H32^{\delta}$ and $Y75^{H}$.

Regarding the system 3) $H32^{\delta}$-Y75 and consistently with system 2), residue $H32^{\delta}$ appears oriented towards the solvent due to the lack of interactions with Y75 (Figure 6.10 e, f). In this case, the hydrogen bond between Y75 and the backbone of G33 is neither possible. Interestingly, the direct consequence of this state, is that the loops A and B are able to maintain open configurations. This suggests that the protonation state of the tyrosine Y75 can be one of the main responsible of the opening or closing of the loops A and B.

Interestingly, the protonation state of system 3) $H32^{\delta}$-Y75, which is the unique that leads to open conformations of loops A and B, is the only able to chelate the heme group. Aiming to assess if the found open states during the MD simulation of system 3) are able to accommodate the heme group, Protein-ligand docking was performed using as host a selected frame corresponding to a properly located histidine that may be able to coordinate the metal center (Figure 6.11). The resulting structure was submitted to 100 ns of MD simulation. Even this work is being currently performed and these results are not yet available, MD simulations suggests two possible binding modes of the heme group at the HasA binding site: one which corresponds with the coordination state found in the X-ray structure (involving $H32^{\delta}$ and Y75) and an

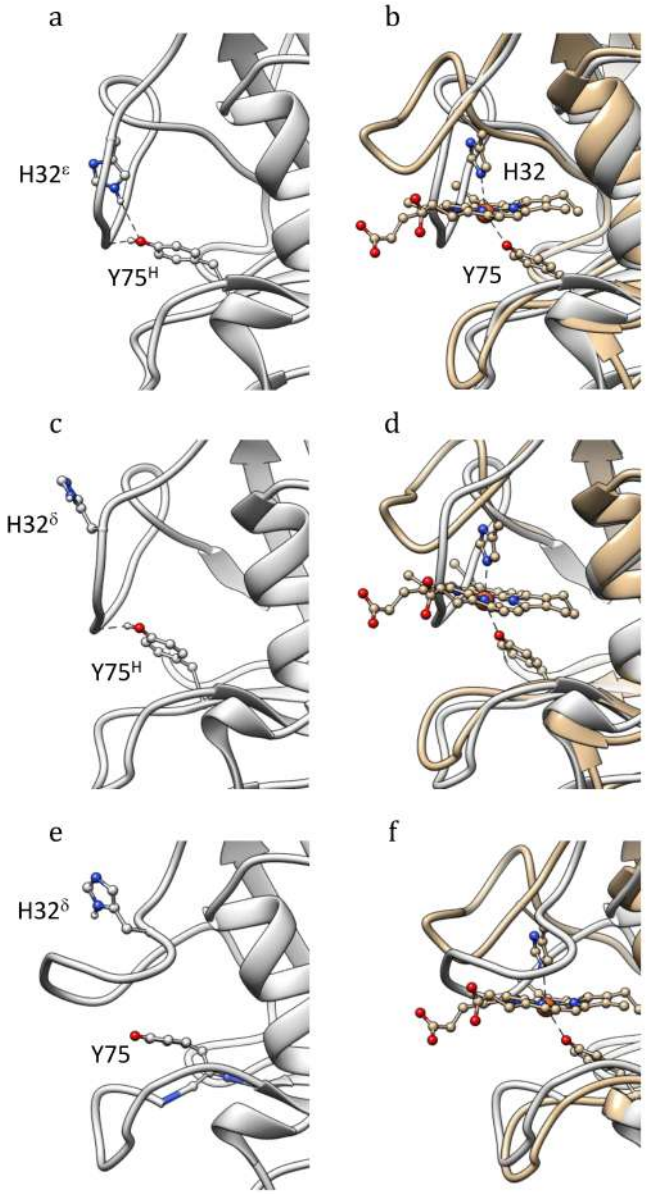


Figure 6.10: Results from 100 ns MD simulation (in grey) and their superimposition with the X-ray structure (in tan) of the HasA protein (PDB code: 1dkh) of representative structures along the simulation of a) H32$^{\varepsilon}$-Y75^{H}, b) H32$^{\delta}$-Y75^{H} and c) H32$^{\delta}$-Y75 systems.

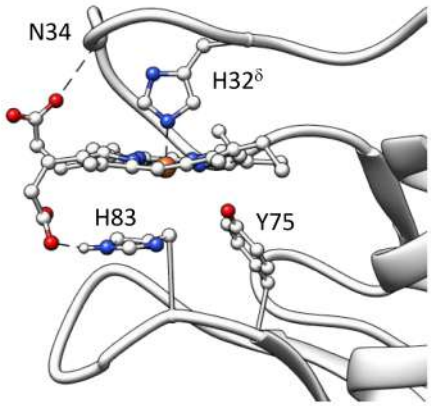


Figure 6.11: Structure resulting from Protein-ligand docking using as host a selected "opened" frame from 100 ns simulation of system 3) H32$^{\delta}$-Y75.

additional one that was not considered before, which involves coordination of the metal center with histidine H83. Even the X-ray structure is consistent only with the first case, the second is supported by previous experimental work in which the binding of the heme was suggested to involve two histidines. [457]

Conclusions

Our results show that the cavity of the *apo* form of the HasA protein can adopts a wide variety of conformations. These appear highly dependent on the protonation state of the residues located at the active site. MD simulations on variants involving protonation states for the coordinating residues H32 and Y75 of the type 1) H32$^{\varepsilon}$-Y75^{H} and 2) H32$^{\delta}$-Y75^{H} do not lead to opened states of the loops responsible of binding the heme moiety (loops A and B). In contrast, in the MD simulation involving the coordination state 3) H32$^{\delta}$-Y75, the system accesses to such opened conformations. Interestingly, this protonation state is the adequate for residues H32 and Y75 coordinate the metal center. This opened state has been successfully used as host to dock the heme group inside the cavity. Additionally, current work suggests an alternative binding mode not observed in the X-ray, involving histidines H32$^{\delta}$ and H83.

This work sustains that the heme recognition process can occur throughout conformational selection. Additionally, results suggest that this is regulated by changes on the protonation state of the coordinating amino acids at the protein active site. The proper combination of Protein-ligand docking and MD simulation techniques results again powerful to elucidate and characterize complex biomolecular process such as, in this case, the conformational regulation of the heme binding in a naturally occurring hemophore.

Chapter 7

Conclusions

The primary goal of this Ph.D. has been to assess if the correct tuning and combination of standard molecular modelling tools (Quantum Mechanics based calculations, Protein-ligand docking, Molecular Dynamics simulations and Quantum Mechanics/Molecular Mechanics) allows the *in silico* driven design of Artificial Metalloenzymes (ArMs). Overall, this objective has been reached by leading to accurate multiscale strategies which not only have allowed rationalizing the catalytic mechanisms of existing ArMs but also designing *de novo* systems.

The discrete objectives reached can be summarize for each chapter:

- Chapter 4 shows that the multiscale strategies allows rationalizing ArMs mechanisms like the hydration of alkenes or the impact of directed evolution of an artificial ATHase. The most important novelty of this work is to confirm that the inclusion of MD simulations in our modelling protocol appeared essential to assess key aspects in the catalytic mechanisms (and more in particular the identification of pre-reactive structures) for the different systems studied.

- Chapter 5 shows the interest of multiscale strategies for the *de novo* enzyme design and re-design. Focusing of including a cooper coordinating unnatural amino acid (UAA) in biological scaffolds, this resulted in the most challenging

study along the thesis and, waiting for experimental validation, this opens the way for the discovery of new catalytically active metallopeptides for copper mediated enantioselective catalysis.

- Chapter 6 finally aimed at better assessing several fundamental aspects that modelling should account for and were characterized along the Ph.D. Using as model systems a recently *de novo* designed heme binding ArM and two natural heme binding proteins, computation was used, first, to understand how active site dynamics can allow to access to pre-catalytic conformations and, second, to gain more insight about how naturally occurring heme binding proteins are able to regulate either the heme binding process and/or the heme mediated catalysis. Again, the combination of Protein-ligand docking and MD simulations appeared indispensable to elucidate the behaviour of the studied systems under different conditions.

For all those systems studied along this thesis for which no crystallographic information was available, the Protein-ligand docking technique appeared essential as it was key to construct the starting point of our model systems. However, refinement via MD simulations was critical to gain more insights about the protein adaptability after the inclusion of the different cofactors and also to have access to pre-catalytic structures that would not have been never accessible via flexible docking simulations.

The full list of studies exposed along this dissertation shows also the challenges that had to be overcome and some of the limitations that we had to confront:

- The first one was related to the lack of crystallographic material for many of the studied Artificial Metalloenzymes, which forced us to generate models without initial knowledge about the type of interactions between the proteins and the catalysts.

- Related with the previous one, came the challenge of dealing with metal ions in non-quantum based approaches. We had to optimize the pipeline for efficient parametrization in force-field based techniques (Protein-ligand docking and MD simulations). Furthermore, we faced the necessity to limit the force-field simulations to discrete states of the catalysis (Michaelis complexes, transition state structures or intermediates) and never consider changes on the chemical state of the cofactor-substrate complexes.

- Another challenge was to generate analytical tools to find pre-reactive structures that could be occupied at room temperature although being minority during the full length of the MD simulation. This point resulted key to rationalize the enantiomeric tendencies observed for the different ArMs studied along this thesis.

- Separately, we had to adapt on the complexity of dealing with a reaction mechanism which involve solvent molecules as reactant. This made extra difficult the characterization of the catalytic pathways as well as working under the QM/MM framework for water molecule activation in solvent exposed binding sites. This actually consist in one of the still remaining limitation in quantum based techniques for simulating catalytic processes: how to deal with water molecule activation when the system is exposed to a network of water molecules.

- Finally, confronting the conformational space of cofactor binding between theoretical calculation and X-ray snapshots is still challenge, especially when loop rearrangements are involved, as was the case in the presented studies.

Despite the numerous challenges, these did not impede us go throughout the different stages of the initial pipeline. We firmly think that the work contained in this dissertation sheds light on the numerous applications resulting from the combination of standard computational tools, specially for studying and designing new ArMs, one

of the most trending and promising fields in protein engineering for the discovery of new-to-Nature reactions.

Appendix A

Chapter 4 - Supplementary information

A.1 The study of the Cu-catalysed hydration of alkenes in water and into the context of an artificial metallohydratase

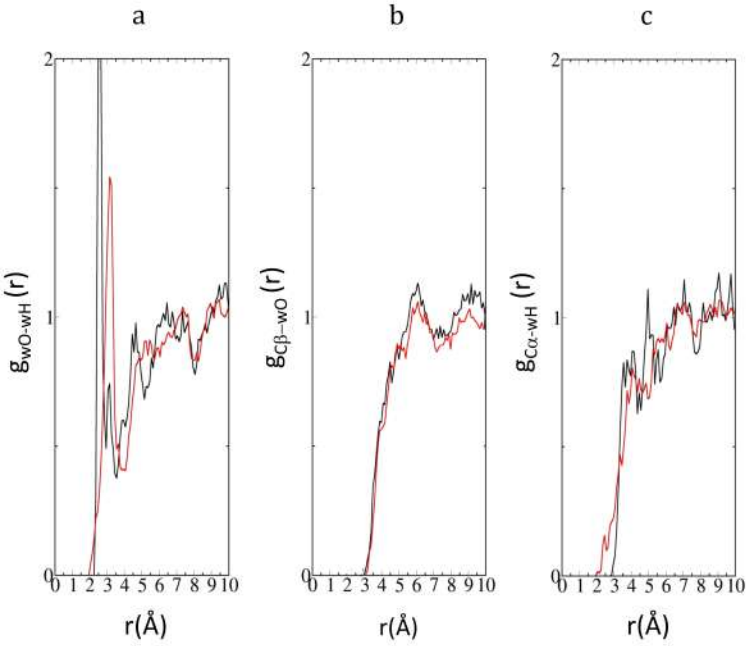


Figure A.1: Comparison between small (red) and big (black) model systems of the radial distribution function for a) the oxygen of the nucleophilic water (w_{NO}) with hydrogens of environmental waters (w_H), b) the beta carbon of the double bond (β) with surrounding water molecules (w_O), skipping the restrained nucleophilic one, and c) the alpha carbon of the double bond (α) and surrounding water molecules (w_H).

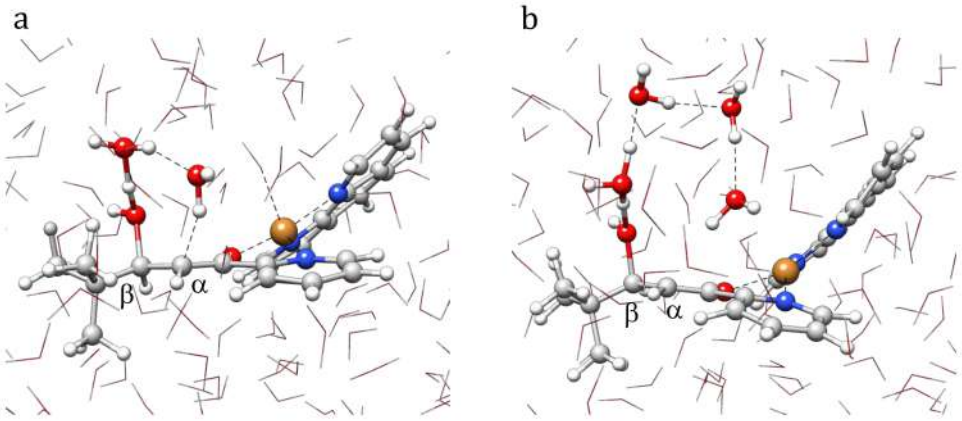


Figure A.2: Examples of two (left) and four (right) member water chains found along 12 ps of restrained AIMD.

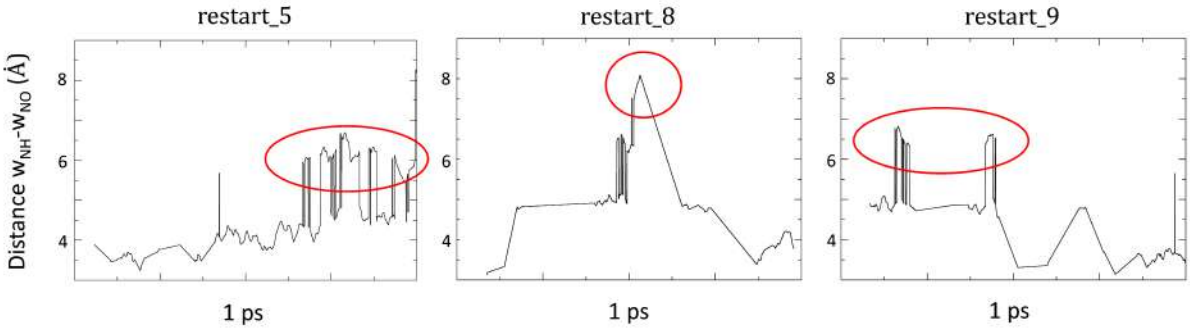


Figure A.3: Distances between the nucleophilic water proton w_{NH} and the nucleophilic water oxygen w_{NO} along 1 ps of unrestrained AIMD. Proton jumping was observed from the starting points selected at 6 (left), 8 (center) and 9 (last) ps of restrained AIMD.

Table A.1: Best docking solutions after the two docking runs for the inclusion of both 1) Phen-Cu(II)-**2b** and 2) Phen-Cu(II)-$2H_2O$ ligands at positions 89 and 89' of LmrR, respectively. The ChemScore scoring function estimates the total free energy (ΔG) resulting from ligand binding by adding to it the magnitude of different particular physical contributions: lipophilicity (S_{lipo}), protein-ligand hydrogen bonding (Shbond), metal binding (S_{metal}) and the loss of conformational entropy of the ligand after binding (H_{rot}). The final ChemScore value (ChemScore column) adds to the total free energy: clash penalty resulting from too close contacts between the ligand and the protein (ΔE_{clash}), internal ligand torsional strain penalty (ΔE_{int}), an internal ligand energy correction term (Intcor) and a protein energy correction term when using flexible sidechains that penalized clashes (S_{prot}).

Scaffold	**Ligand**	**ChemScore**	**ΔG**	S_{hbond}	S_{metal}	S_{lipo}	H_{rot}	ΔE_{clash}	ΔE_{int}	ΔE_{cov}	**Intcor**	S_{prot}
LmrR_M89C	1	50.06	-38.75	0.00	0.00	310.35	1.19	-13.33	0.35	0.40	1.44	1.26
LmrR_M89C_1	2	37.19	-28.23	0.00	0.00	220.55	1.19	-13.32	0.41	0.90	1.22	3.05

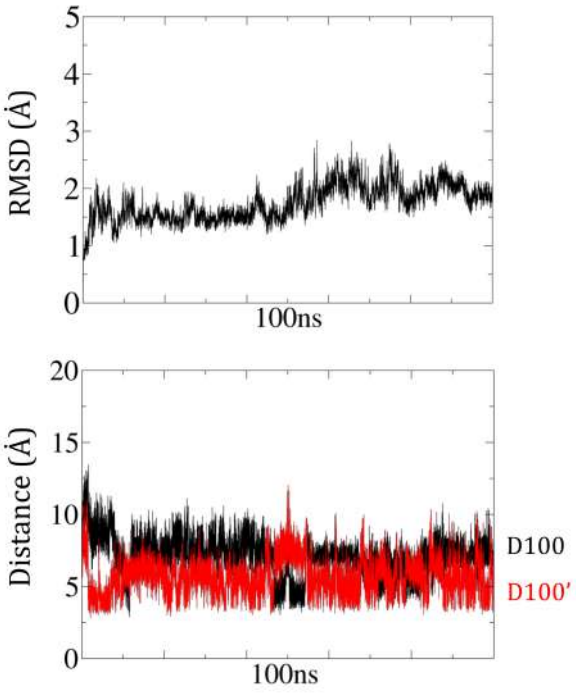
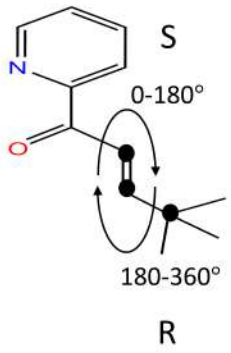


Figure A.4: RMSD (top) and distance (bottom) in $\mathring{A}$ between the substrate double bond and D100/D100' residues along 100 ns MD simulation.



Pro-S face	Pro-R face
-	D100' – 2721 D100 – 1 E104 – 26

Figure A.5: Number of pre-catalytic structures found among 25000 frames selected from 100 ns MD simulation.

A.2 An artificial enantioselective imine reductase

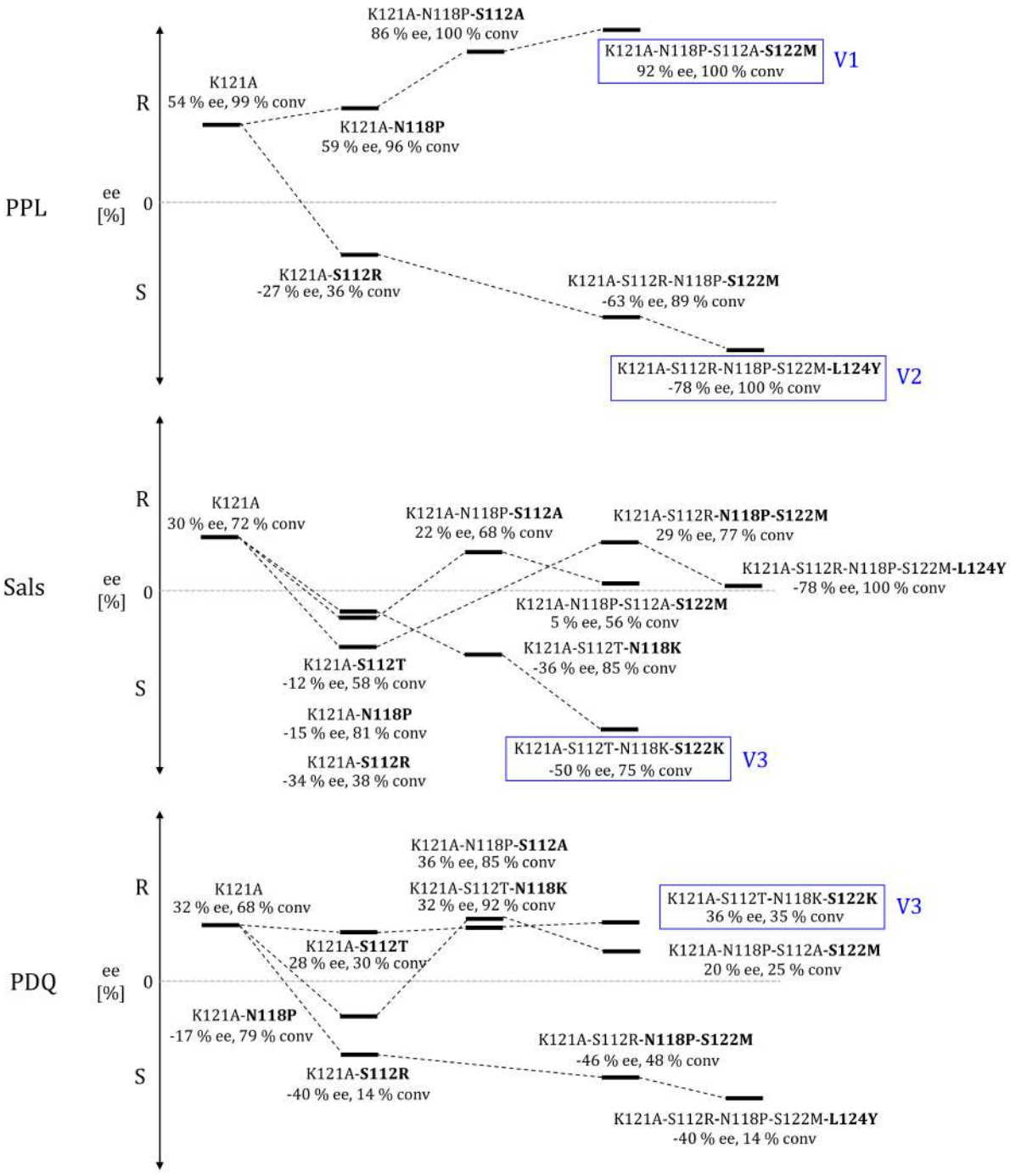


Figure A.6: Directed evolution tree to afford the (R)- or (S)-selective ATHase for PPL, salsolidine and PDQ substrates. The variants selected for the computational study are highlighted in blue.

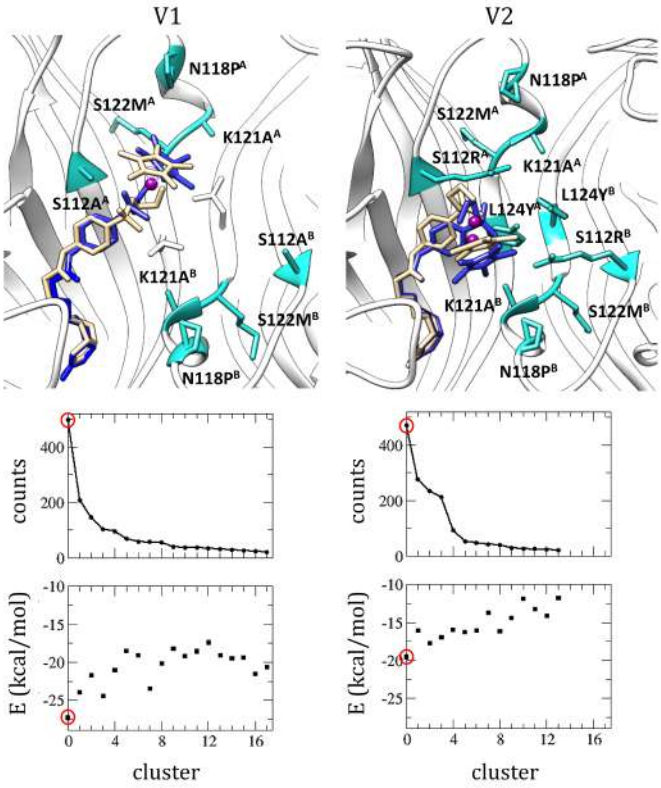


Figure A.7: Superimposition of the Ir cofactors between the selected clusters resulting from the metadynamics for variants V1 (left) and V2 (right), in blue, and their corresponding X-ray structure, in tan.

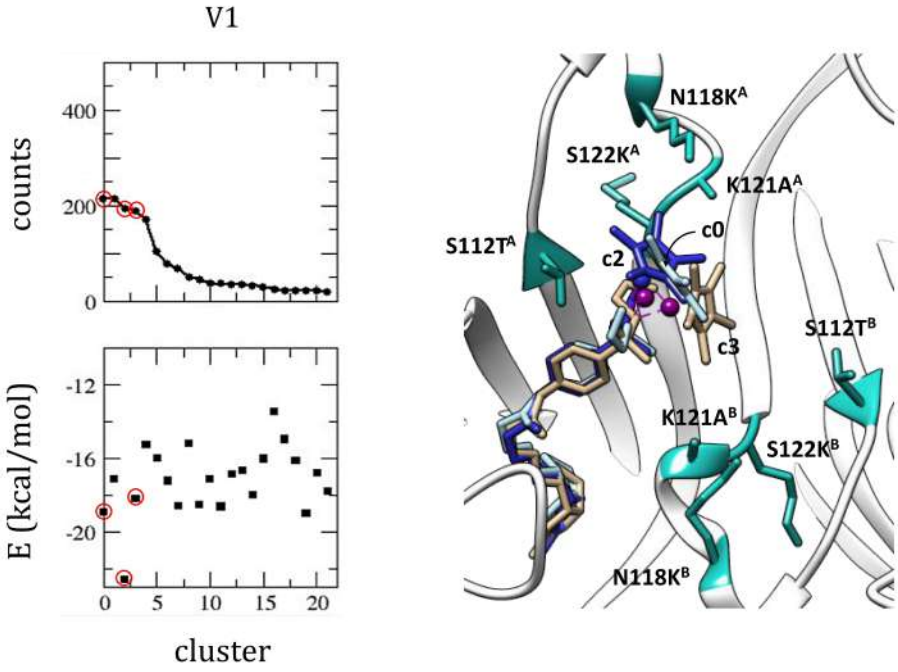


Figure A.8: Selected clusters for the variant V1, for which no crystallographic material was described.

Appendix B

Chapter 5 - Supplementary information

B.1 An artificial metallohydratase including the non-canonical amino acid Ala-2,2'-bipyridine

Table B.1: Selected docking solutions for the LmrR and its variants bound to BpyA-Cu(II)-1a complex. The ChemScore scoring function estimates the total free energy (ΔG) resulting from ligand binding by adding to it the magnitude of different particular physical contributions: lipophilicity (S_{lipo}), protein-ligand hydrogen bonding (Shbond), metal binding (S_{metal}) and the loss of conformational entropy of the ligand after binding (H_{rot}). The final ChemScore value (ChemScore column) adds to the total free energy: clash penalty resulting from too close contacts between the ligand and the protein (ΔE_{clash}), internal ligand torsional strain penalty (ΔE_{int}), an internal ligand energy correction term (Intcor) and a protein energy correction term when using flexible sidechains that penalized clashes (S_{prot}).

Variant	**Sys**	**Cof(s)**	**ChemScore**	**ΔG**	S_{hbond}	S_{metal}	S_{lipo}	H_{rot}	ΔE_{clash}	ΔE_{int}	ΔE_{cov}	**Intcor**	S_{prot}
M89X	1	–	33.76	-15.85	0.00	0.00	88.62	0.00	-19.20	0.00	0.08	0.00	1.20
		1	33.79	-16.06	0.00	0.00	90.46	0.00	-19.04	0.00	0.09	0.00	1.22
	2	–	47.39	-30.18	0.00	0.00	233.01	1.00	-18.83	0.15	0.26	0.00	1.21
		1	33.79	-16.03	0.00	0.00	90.19	0.00	-19.05	0.00	0.07	0.00	1.22
		2	42.07	-27.69	0.00	0.00	211.67	1.00	-17.22	0.04	1.57	0.00	1.23
M89X_V15E	**2**	–	44.40	-30.23	0.00	0.00	233.38	1.00	-17.96	0.21	0.41	0.00	3.16
		1	32.2	-14.47	0.00	0.00	76.87	0.00	-19.02	0.00	0.07	0.00	1.22
		2	41.01	-27.03	0.00	0.00	206.05	1.00	-16.17	0.03	0.91	0.00	1.24
M89X_W96E	**2**	–	40.00	-22.76	0.00	0.00	169.54	1.00	-18.73	0.03	0.26	0.00	1.20
		1	33.85	-16.26	0.00	0.00	92.14	0.00	-18.98	0.00	0.15	0.00	1.24
		2	39.48	-23.20	0.00	0.00	173.30	1.00	-18.95	0.00	1.44	0.00	1.22
M89X_D100E	**2**	–	47.58	-30.38	0.00	0.00	234.71	1.00	-18.83	0.17	0.26	0.00	1.20
		1	33.95	-16.39	0.00	0.00	93.23	0.00	-18.94	0.00	0.17	0.00	1.21
		2	42.36	-27.93	0.00	0.00	213.73	1.00	-17.26	0.04	1.58	0.00	1.21

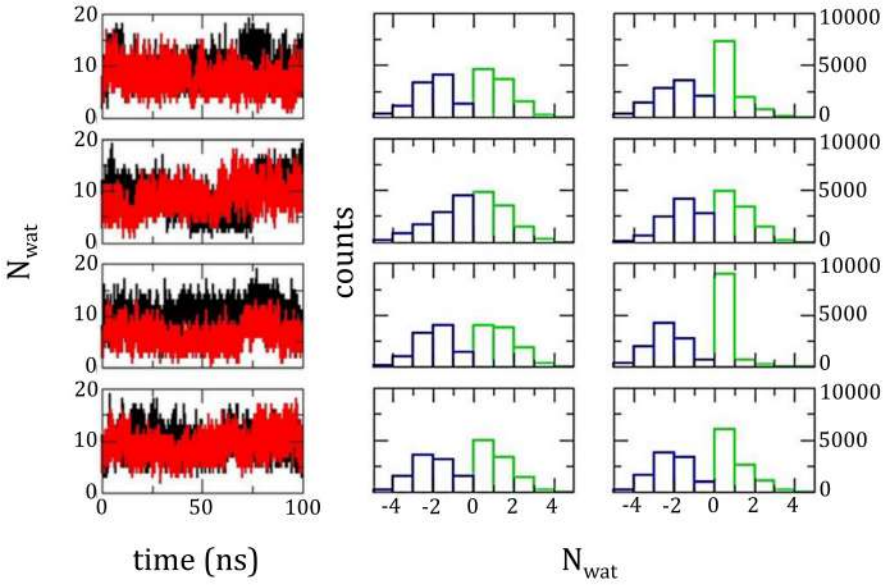


Figure B.1: Data of the MD simulations for (top to bottom) LmrR_M89X, LmrR_M89X_D100E, LmrR_M89X_V15E, LmrR_M89X_W96E. *Left histogram*: number of water molecules (N_{wat}) within 5 $\AA$ from the β carbon of substrate 1a. *Middle and right histograms*: pre-catalytic N_{wat} with respect to the β carbon of the substrates 1a and 1a', respectively, blue for the pro-S face and green for the pro-R face.

Table B.2: For each variant, the table shows the number of times a pre-reactive conformation involving an Asp or Glu residue is observed. 10.000 frames from 100 ns of MD simulations were analysed.

	BpyA-Cu(II)-1a		**BpyA-Cu(II)-1a'**	
Variant	**Pro-*R***	**Pro-*S***	**Pro-*R***	**Pro-*S***
M89X	D100'-866 E104'-1 E107'-6	E94-34 D100'-7 E104'-1	D100-100 E104-9 E107-14	–
M89X_D100E	E100-45 E100'-257 E104'-16 E107'-193	E87-45 E100-5 E100'-30	E100-366 E104-13 E100'-13	E100-34 E94'-1 E100'-1
M89X_V15E	E15-904 D100'-1235 E107'-2	E15-503 D100-1 D100'-215	E15'-263	D100-4
M89X_W96E	E96-384 D100-19 E96'-293 D100'-793 E107'-3	E87-19 E96-908 D100-58 E96'-2744 D100'-18	E7-32 E96-314 D100-80 E104-64 E107-13 E97'-32	E7-3 E96-1650 D100-384 E104-4 E97'-1238

Table B.3: Number of pre-reactive frames with a distance less than 5 $\AA$ between the β carbon of the substrate and the oxigens of the D/E residues . 10.000 frames from 100 ns of MD simulations were analysed.

	BpyA-Cu(II)-1a		**BpyA-Cu(II)-1a'**	
Variant	**Pro-*R***	**Pro-*S***	**Pro-*R***	**Pro-*S***
M89X	113	1	123	0
M89X_D100E	44	12	51	5
M89X_V15E	541	217	97	0
M89X_W96E	739	1528	172	1300

Table B.4: Docking predicted distances (Å) among the mutated residues and the β-carbon of the double bond of both substrates.

Variant	**Residue**	**O-Cβ**	**O-Cβ'**
M89X	**D100**	5.72	6.04
	D100'	6.22	8.62
M89X_V15E	**V15E**	3.76	6.85
	V15E'	7.34	3.74
M89X_W96E	W96E	3.23	5.42
	W96E'	3.91	5.45
M89X_D100E	D100E	5.51	6.51
	D100E'	7.62	9.87
M89X_M8E	M8E	8.75	3.95
	M8E'	8.67	10.21
M89X_A92E	A92E	15.42	16.92
	A92E'	17.66	13.63
M89X_S95E	S95E	9.79	13.73
	S95E'	13.18	12.21
M89X_Q12E	Q12E	8.34	7.12
	Q12E'	7.57	8.26

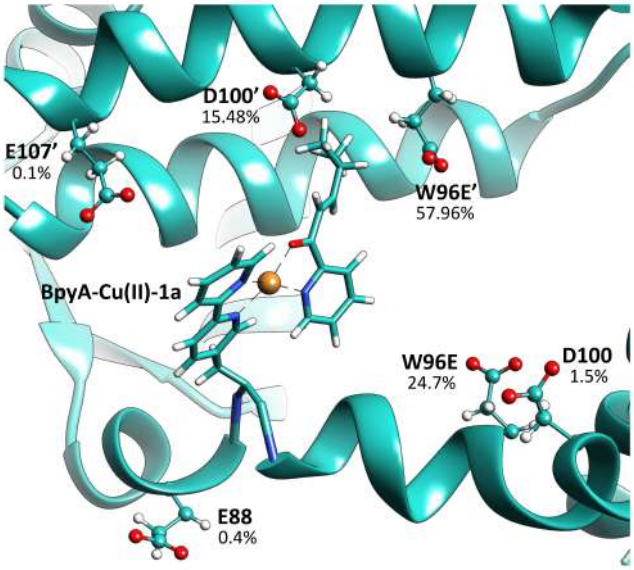


Figure B.2: Representation of the contribution of each residue to form pre-catalytic configurations in the variant LmrR_M89X_W96E.

Appendix C

Chapter 6 - Supplementary information

C.1 An artificial heme enzyme for enantioselective cyclopropanation reactions

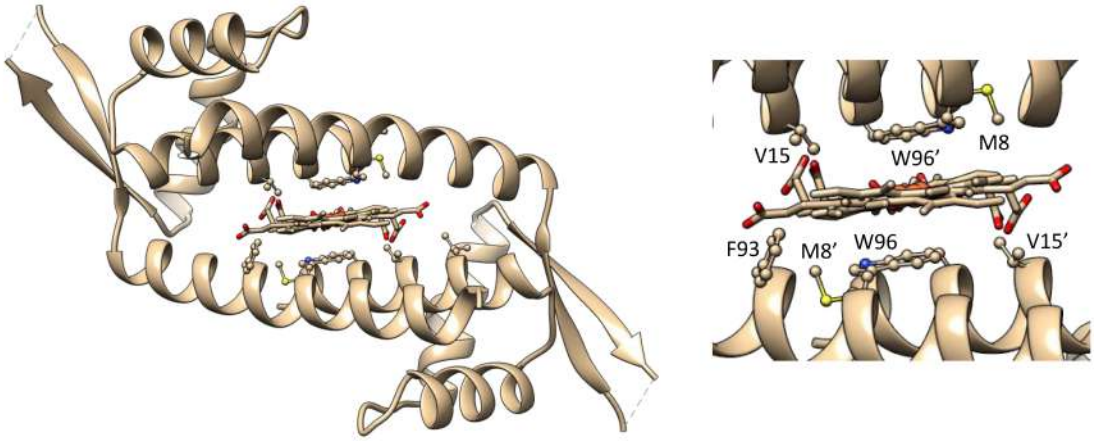


Figure C.1: Solved crystal structure of the LmrR bound to the heme group (PDB code: 6UUF).

Table C.1: Experimental results of the enantioselective cyclopropanation reaction catalysed by the LmrR⊂heme system and its variants. [171]

entry	LmrR	1[c]	3	Yield %	TTN	3/4	ee %[b]
1	-	**1a**	**3a**	5 ± 2	51	0.8	-
2	LmrR	**1a**	**3a**	25 ± 11	247	11	17 ± 5
3	LmrR _F93A	**1a**	**3a**	23 ± 2	232	9	11 ± 1
4	LmrR _D100A	**1a**	**3a**	38 ± 8	375	20	24 ± 5
5	LmrR _W96A	**1a**	**3a**	28 ± 13	276	8	< 5
6	LmrR _V15A	**1a**	**3a**	1.5 ± 0.5	15	3	17 ± 1
7	LmrR _M8A	**1a**	**3a**	36 ± 13	359	15	44 ± 12
8[c]	-	**1a**	**3a**	5 ± 2	51	0.8	-
9[c]	LmrR _M8A	**1a**	**3a**	45 ± 9	449	6	51 ± 14
10[c]	-	**1b**	**3b**	6 ± 0	59	n.d	-
11[c]	LmrR _M8A	**1b**	**3b**	39 ± 13	391	n.d	38 ± 5[d]
12[c]	-	**1c**	**3c**	1 ± 0	12	0.2	-
13[c]	LmrR _M8A	**1c**	**3c**	35 ± 13	351	3	25 ± 5

[a] Conditions: 1 (30 mM), 2 (10 mM), hemin (1 mol%; 10 μM), LmrR_X (1.1 mol%; 11 μM) in 50 mM phosphate buffer (pH 8.0), under Ar, at 4°C for 18h; Results are the average of at least two independent experiments, both carried out in duplicate. [b] of the trans product; trans:cis > 85:15. [c] pH 7.0. [d] of the 1*R*, 2*R* enantiomer.

Table C.2: Selected docking solutions for the LmrR and its variants bound to the heme group (**5**), the heme-carbene complex (**6**) and the heme-TS complex corresponding to the 1R,2R enantiomeric product (**7**). The ChemScore scoring function estimates the total free energy (ΔG) resulting from ligand binding by adding to it the magnitude of different particular physical contributions: lipophilicity (S_{lipo}), protein-ligand hydrogen bonding (S_{hbond}), metal binding (S_{metal}) and the loss of conformational entropy of the ligand after binding (H_{rot}). The final ChemScore value (ChemScore column) adds to the total free energy: clash penalty resulting from too close contacts between the ligand and the protein (ΔE_{clash}), internal ligand torsional strain penalty (ΔE_{int}), an internal ligand energy correction term (Intcor) and a protein energy correction term when using flexible sidechains that penalized clashes (S_{prot}).

Entry	**Final complexes**	**ChemScore**	**ΔG**	S_{hbond}	S_{metal}	S_{lipo}	H_{rot}	ΔE_{clash}	ΔE_{int}	**Intcor**	S_{prot}
1	**LmrR⊂5_1**	55.06	-55.54	1.96	0.00	433.00	2.79	0.12	0.35	0.06	0.00
2	**LmrR⊂5_2**	54.90	-55.34	1.97	0.00	431.19	2.79	0.28	0.16	0.06	0.00
3	**LmrR⊂5_3**	51.34	-52.16	0.97	0.00	432.42	2.79	0.21	0.61	0.06	0.00
4	**LmrR⊂5_4**	51.22	-52.88	0.97	0.00	438.51	2.79	0.81	0.84	0.06	0.00
5	**LmrR⊂5_W96A_1**	33.58	-34.76	0.99	0.00	283.15	2.79	0.77	0.42	0.07	0.00
6	**LmrR⊂5_W96A_2**	33.05	-33.67	0.94	0.00	272.57	2.79	0.51	0.09	0.07	0.00
7	**LmrR⊂5_W96A_3**	32.79	-33.67	0.94	0.00	272.57	2.79	0.49	0.10	0.06	0.00
8	**LmrR⊂6_1**	52.62	-57.09	1.77	0.00	459.97	3.18	3.82	0.65	0.20	0.00
9	**LmrR⊂6_2**	48.45	-50.17	0.97	0.00	423.85	3.18	1.14	0.59	0.20	0.00
10	**LmrR⊂6_3**	48.26	-53.40	1.78	0.00	428.31	3.18	2.08	1.17	0.20	1.90
11	**LmrR⊂6_4**	47.22	-53.12	0.96	0.00	449.37	3.18	3.29	0.74	0.20	1.87
12	**LmrR⊂7**	48.83	-61.49	0.98	0.00	525.40	3.41	7.41	3.68	0.71	1.57
13	**LmrR⊂7_M8A**	45.67	-56.65	0.99	0.00	483.77	3.41	7.87	1.90	0.79	1.22

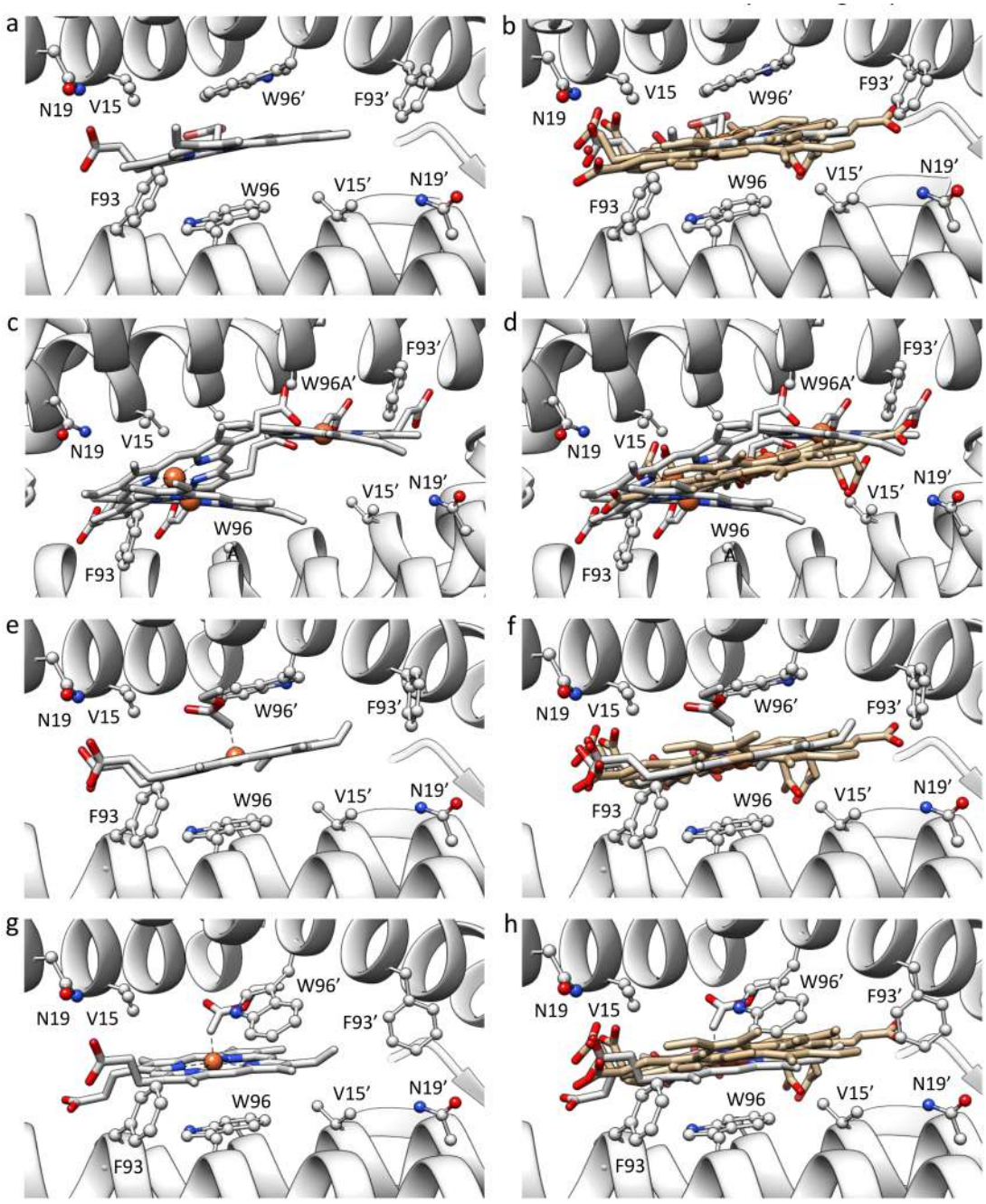


Figure C.2: The column in the left corresponds with structures resulting from docking and the column on the right, these superimposed with the X-ray structure. The image shows selected docking solutions for: the heme group bound to a) b) LmrR and c) d) LmrR_W96A; e) f) the LmrR bound to the heme-carbene complex with the W96' pointing towards the cavity maintaining a "closed" orientation similar to the X-ray structure; and g) h) the LmrR bound to the heme-carbene complex with the W96' flipped towards the solvent leading to a "opened" conformation.

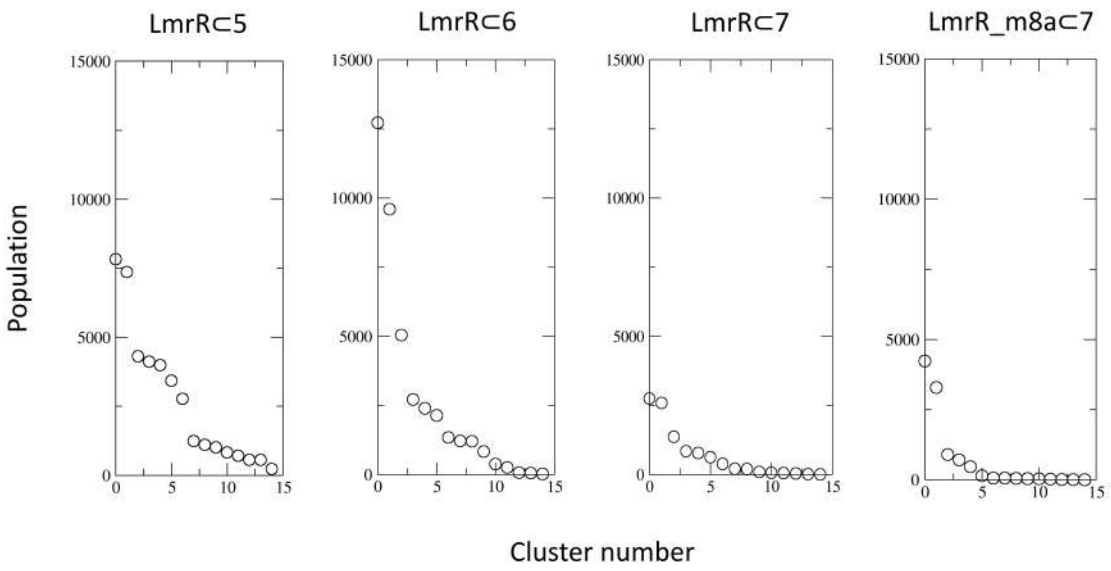


Figure C.3: Cluster analysis of the MD simulations for the LmrR bound to heme (**5**) and its adducts with the carbene substrate (**6**) and with transition state of the cyclopropanation reaction (**7**). For the last, both the WT LmrR and the variant LmrR_M8A were assessed.

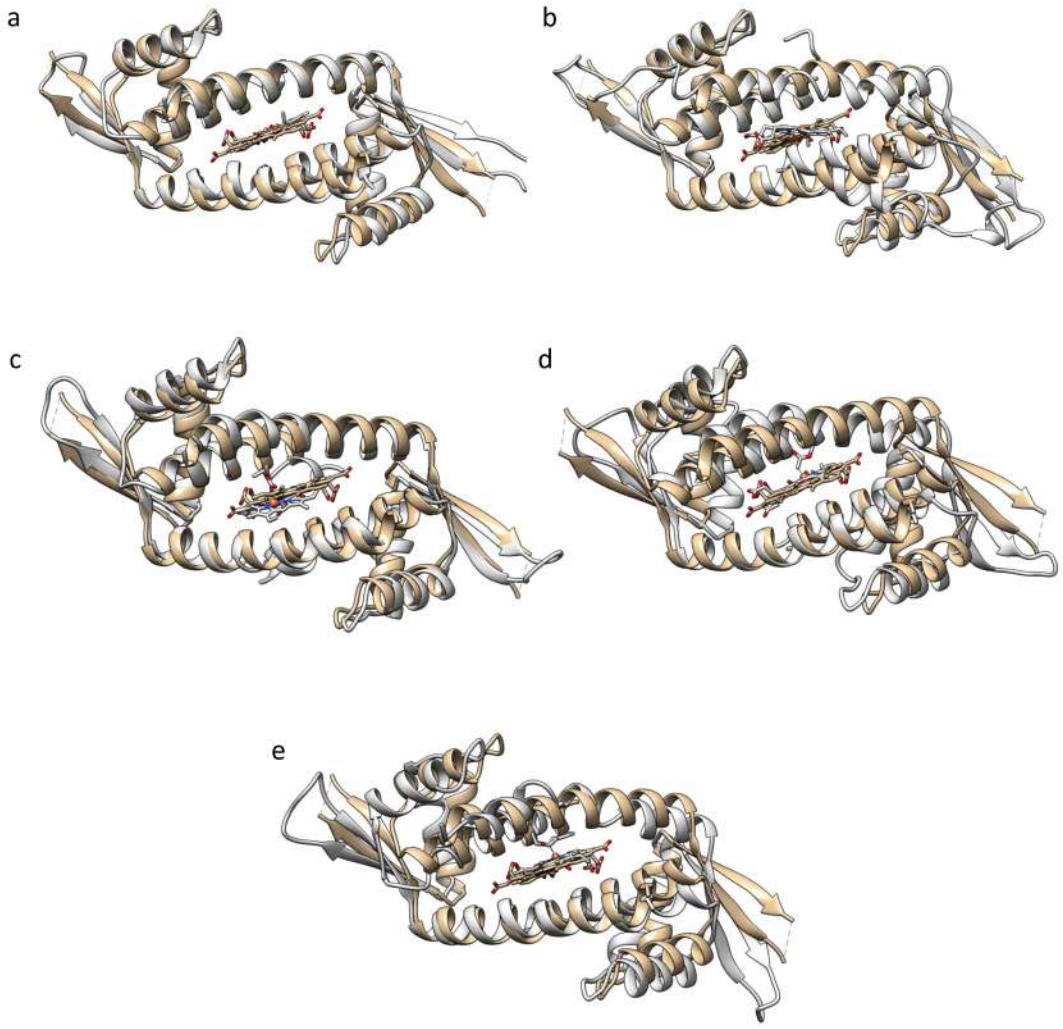


Figure C.4: Superimposition of the X-ray structure of the LmrR⊂heme complex (PDQ code: 6UUF - silver) and representative structures resulting from MD simulations (gold) of: the LmrR-heme system (a) cluster 3, representative frame of a non-catalytic closed structure and (b) cluster 2, representative frame of an open structure accessible to substrates. Data collected from 400 ns MD simulation), the LmrR-heme-carbene system (c) cluster 0, representative frame of a closed structure in which the heme-carbene complex is displaced towards the solvent and d) cluster 2, an open conformation in which the heme-carbene complex fits at the active site of the LmrR protein. Data collected from 400 ns MD simulation), and e) the LmrR protein bound to the transition state for the cyclopropanation reaction (involving the heme bound to the carbene and styrene substrates), which promotes the opening of the active site and allows the complex to be located at the dimer interface.

C.2 Structure-function relationship in naturally ocurring hemophores

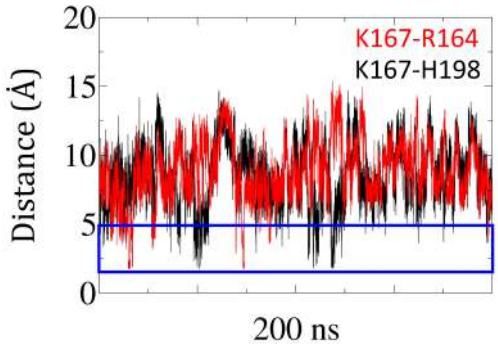


Figure C.5: Distances analaysis corresponding to 200 ns of MD simulation of the *apo* form of the LCP protein in its closed state (5o1m) and with deprotonated K167. Graph shows the distances between the nitrogen of K167 and the corresponding hydrogen of R164 (labelled in red) and H198 (labelled in black).

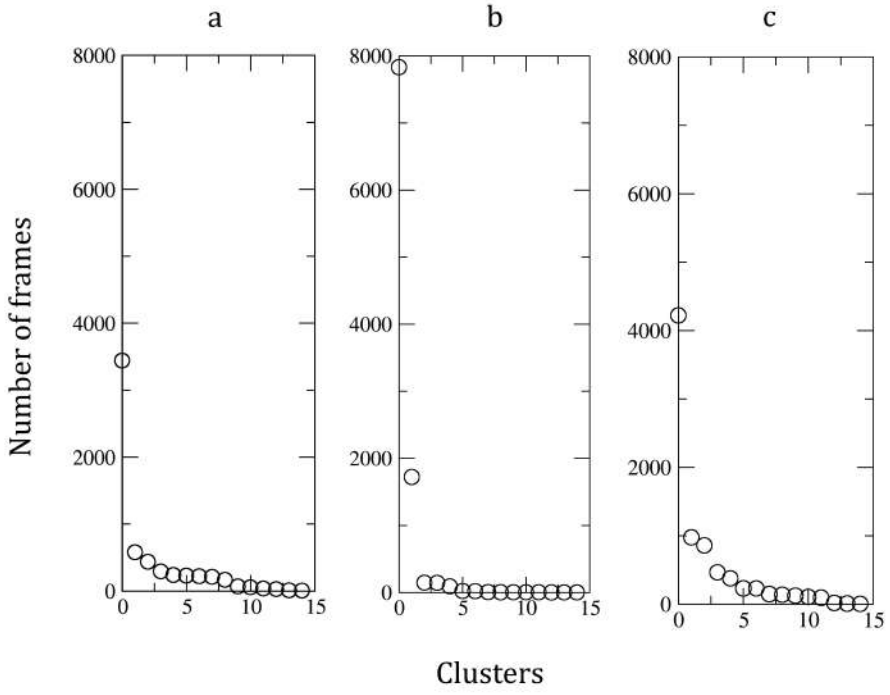


Figure C.6: Cluster analysis of the MD simulations for the systems a. H32$^{\varepsilon}$-Y75^{H}, b. H32$^{\delta}$-Y75^{H} and c. H32$^{\delta}$-Y75 of the HasA protein.

Appendix D

List of publications

D.1 Articles

1. L. Villarino, K. Splan, E. Reddem, C. Gutierrez de Souza, L. Alonso-Cotchico, A. Lledos, J.-D. Maréchal, A.-M. Thunnissen, and G. Roelfes, "An artificial heme enzyme for cyclopropanation reactions" *Angewandte Chemie - International Edition*, 10.1002/anie.201802946, 2018.

2. M. Hestericová, T. Heinisch, L. Alonso-Cotchico, J. D. Maréchal, P. Vidossich, and T. R. Ward, "Directed evolution of an artificial imine reductase," *Angewandte Chemie - International Edition*, vol. 57, no. 7, pp. 1863–1868, 2018.

3. I. Drienovská, L. Alonso-Cotchico, P. Vidossich, A. Lledós, J.-D. Maréchal, and G. Roelfes, "Design of an enantioselective artificial metallo-hydratase enzyme containing an unnatural metal-binding amino acid" *Chemical Science*, vol. 8, no. 10, pp. 7228–7235, 2017.

4. W. Ghattas, J.-P. Mahy, R. Ricoux, J.-D. Maréchal, A. Urvoas, L. Alonso-Cotchico, and M. Rousseau, "Artificial metalloenzymes with the neocarzinostatinscaffold : toward a biocatalyst for the diels-alder cyclization reaction." *ChemBioChem*, vol. 17, no. 5, pp. 433–440, 2015.

D.2 Reviews

1. VM. Robles, E. Ortega-Carrasco, L. Alonso-Cotchico, J. Rodriguez-Guerra, A. Lledós, and J.-D. Maréchal, "Toward the computational design of artificial metalloenzymes: From protein-ligand docking to multiscale approaches." *ACS Catalysis*, vol. 5, no. 4, pp. 2469-2480, 2015

D.3 Book chapters

1. J. Rodríguez-Guerra, L. Alonso-Cotchico, G. Sciortino, A. Lledós, and J.-D. Marechal, "Computational studies of artificial metalloenzymes: From methods and models to design and optimization," in *Artificial metalloenzymes and metalloDNAzymes in catalysis: From design to application*, pp. 99–136, Wiley-VCH Verlag GmbH & Co. KGaA Weinheim, Germany, 2018.

2. L. Alonso-Cotchico, J. Rodríguez-Guerra, A. Lledós, and J.-D. Maréchal, "Enzyme design," in *Simulating Enzyme Reactivity*, pp. 481–521, The Royal Society of Chemistry, 2016.

Bibliography

[1] M. Arroyo, C. Acebal, and I. De la Mata, "Biocatálisis y biotecnología," *Arbor*, vol. 190, no. 768, p. a156, 2014.

[2] B. M. Nestl, S. C. Hammer, B. A. Nebel, and B. Hauer, "New generation of biocatalysts for organic synthesis," *Angewandte Chemie - International Edition*, vol. 53, no. 12, pp. 3070–3095, 2014.

[3] M. T. Reetz, "Biocatalysis in organic chemistry and biotechnology : Past , present , and future," *Journal of the Chemical Society*, vol. 135, pp. 12480–12496, 2013.

[4] U. T. Bornscheuer and R. J. Kazlauskas, "Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways.," *Angewandte Chemie - International Edition*, vol. 43, no. 45, pp. 6032–6040, 2004.

[5] B. G. Davis and V. Boyer, "Biocatalysis and enzymes in organic synthesis.," *Natural Product Reports*, vol. 18, no. 6, pp. 618–640, 2001.

[6] R. A. Sheldon and S. Van Pelt, "Enzyme immobilisation in biocatalysis: why, what and how," *Chemical Society Reviews*, vol. 42, no. 42, pp. 6223–6235, 2013.

[7] P.-S. Huang, S. E. Boyken, and D. Baker, "The coming of age of de novo protein design," *Nature*, vol. 537, no. 7620, pp. 320–327, 2016.

[8] P. B. Harbury, "High-resolution protein design with backbone freedom," *Science*, vol. 282, no. 5393, pp. 1462–1467, 1998.

[9] A. R. Thomson, C. W. Wood, A. J. Burton, G. J. Bartlett, R. B. Sessions, R. L. Brady, and D. N. Woolfson, "Computational design of water-soluble α-helical barrels," *Science*, vol. 346, no. 6208, pp. 485–488, 2014.

[10] G. Grigoryan and W. F. Degrado, "Probing designability via a generalized model of helical bundle geometry," *Journal of Molecular Biology*, vol. 405, no. 4, pp. 1079–1100, 2011.

[11] P. S. Huang, G. Oberdorfer, C. Xu, X. Y. Pei, B. L. Nannenga, J. M. Rogers, F. DiMaio, T. Gonen, B. Luisi, and D. Baker, "High thermodynamic stability of parametrically designed helical bundles," *Science*, vol. 346, no. 6208, pp. 481–485, 2014.

[12] N. H. Joh, T. Wang, M. P. Bhate, R. Acharya, Y. Wu, M. Grabe, M. Hong, G. Grigoryan, and W. F. DeGrado, "De novo design of a transmembrane zn(2+)-transporting four-helix bundle.," *Science*, vol. 346, no. 6216, pp. 1520–4, 2014.

[13] L. Regan and W. DeGrado, "Characterization of a helical protein designed from first principles," *Science*, vol. 241, no. 4868, pp. 976–978, 1988.

[14] Y.-R. Lin, N. Koga, R. Tatsumi-Koga, G. Liu, A. F. Clouser, G. T. Montelione, and D. Baker, "Control over overall shape and size in de novo designed proteins," *Proceedings of the National Academy of Sciences*, vol. 112, no. 40, pp. E5478–E5485, 2015.

[15] N. Koga, R. Tatsumi-Koga, G. Liu, R. Xiao, T. B. Acton, G. T. Montelione, and D. Baker, "Principles for designing ideal protein structures," *Nature*, vol. 491, no. 7423, pp. 222–227, 2012.

[16] K. T. Simons, C. Kooperberg, E. Huang, and D. Baker, "Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions," *Journal of Molecular Biology*, vol. 268, no. 1, pp. 209–225, 1997.

[17] B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, and D. Baker, "Design of a novel globular protein fold with atomic-level accuracy," *Science*, vol. 302, no. 5649, pp. 1364–1368, 2003.

[18] P.-S. Huang, Y.-E. A. Ban, F. Richter, I. Andre, R. Vernon, W. R. Schief, and D. Baker, "Rosettaremodel: a generalized framework for flexible backbone protein design.," *PloS one*, vol. 6, no. 8, p. e24109, 2011.

[19] S. T. R. Walsh, H. Cheng, J. W. Bryson, H. Roder, and W. F. DeGrado, "Solution structure and dynamics of a de novo designed three-helix bundle protein," *Proceedings of the National Academy of Sciences*, vol. 96, no. 10, pp. 5486–5491, 1999.

[20] J. M. Fletcher, A. L. Boyle, M. Bruning, G. J. Bartlett, T. L. Vincent, N. R. Zaccai, C. T. Armstrong, E. H. C. Bromley, P. J. Booth, R. L. Brady, A. R. Thomson, and D. N. Woolfson, "A basis set of ¡i¿de novo¡/i¿ coiled-coil peptide oligomers for rational protein design and synthetic biology," *ACS Synthetic Biology*, vol. 1, no. 6, pp. 240–250, 2012.

[21] N. R. Zaccai, B. Chi, A. R. Thomson, A. L. Boyle, G. J. Bartlett, M. Bruning, N. Linden, R. B. Sessions, P. J. Booth, R. L. Brady, and D. N. Woolfson, "A de novo peptide hexamer with a mutable channel," *Nature Chemical Biology*, vol. 7, no. 12, pp. 935–941, 2011.

[22] D. Eisenberg, W. Wilcox, S. M. Eshita, P. M. Pryciak, S. P. Ho, and W. F. Degrado, "The design, synthesis, and crystallization of an alpha-helical peptide,"

Proteins: Structure, Function, and Bioinformatics, vol. 1, no. 1, pp. 16–22, 1986.

[23] A. E. Keating, V. N. Malashkevich, B. Tidor, and P. S. Kim, "Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils," *Proceedings of the National Academy of Sciences*, vol. 98, no. 26, pp. 14825–14830, 2001.

[24] G. Bhardwaj, V. K. Mulligan, C. D. Bahl, J. M. Gilmore, P. J. Harvey, O. Cheneval, G. W. Buchko, S. V. Pulavarti, Q. Kaas, A. Eletsky, P. S. Huang, W. A. Johnsen, P. J. Greisen, G. J. Rocklin, Y. Song, T. W. Linsky, A. Watkins, S. A. Rettie, X. Xu, L. P. Carter, R. Bonneau, J. M. Olson, E. Coutsias, C. E. Correnti, T. Szyperski, D. J. Craik, and D. Baker, "Accurate de novo design of hyperstable constrained peptides," *Nature*, vol. 538, no. 7625, pp. 329–335, 2016.

[25] B. Webb and A. Sali, *Comparative protein structure modeling using MODELLER*, vol. 15. John Wiley & Sons, 2014.

[26] C. A. Rohl, C. E. M. Strauss, K. M. S. Misura, and D. Baker, "Protein structure prediction using rosetta," *Methods in Enzymology*, vol. 383, no. 2003, pp. 66–93, 2004.

[27] K. W. Kaufmann, G. H. Lemmon, S. L. Deluca, J. H. Sheehan, and J. Meiler, "Practically useful: What the r osetta protein modeling suite can do for you," *Biochemistry-US*, vol. 49, no. 14, pp. 2987–2998, 2010.

[28] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. E. Meyer Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, "The protein data bank: A computer-base archival file for macromolecular structures," *Journal of Molecular Biology*, vol. 112, p. 535, 1977.

[29] D. Hilvert, "Design of protein catalysts," *Annual Review of Biochemistry*, vol. 82, no. 1, pp. 447–470, 2013.

[30] H. Kries, R. Blomberg, and D. Hilvert, "De novo enzymes by computational design," *Current Opinion in Chemical Biology*, vol. 17, no. 2, pp. 221–228, 2013.

[31] G. Kiss, Celebi-Ölçüm, R. Moretti, D. Baker, and K. N. Houk, "Computational enzyme design," *Angewandte Chemie - International Edition*, vol. 52, no. 22, pp. 5700–5725, 2013.

[32] D. Baker, "An exciting but challenging road ahead for computational enzyme design," *Protein Science*, vol. 19, no. 10, pp. 1817–1819, 2010.

[33] G. Kiss, D. Röthlisberger, D. Baker, and K. N. Houk, "Evaluation and ranking of enzyme designs.," *Protein Science*, vol. 19, no. 9, pp. 1760–73, 2010.

[34] X. Garrabou, B. I. M. Wicky, and D. Hilvert, "Fast knoevenagel condensations catalyzed by an artificial schiff-base-forming enzyme," *Journal of the American Chemical Society*, vol. 138, no. 22, pp. 6972–6974, 2016.

[35] L. Giger, S. Caner, R. Obexer, P. Kast, D. Baker, N. Ban, and D. Hilvert, "Evolution of a designed retro-aldolase leads to complete active site remodeling," *Nature Chemical Biology*, vol. 9, no. 8, pp. 494–498, 2013.

[36] J. Bos, W. R. Browne, A. J. Driessen, and G. Roelfes, "Supramolecular assembly of artificial metalloenzymes based on the dimeric protein lmrr as promiscuous scaffold," *Journal of the American Chemical Society*, vol. 137, no. 31, pp. 9796–9799, 2015.

[37] A. Pavelka, E. Chovancova, and J. Damborsky, "Hotspot wizard: A web server for identification of hot spots in protein engineering," *Nucleic Acids Research*, vol. 37, no. SUPPL. 2, 2009.

[38] M. Pavlova, M. Klvana, Z. Prokop, R. Chaloupkova, P. Banas, M. Otyepka, R. C. Wade, M. Tsuda, Y. Nagata, and J. Damborsky, "Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate," *Nature Chemical Biology*, vol. 5, no. 10, pp. 727–733, 2009.

[39] R. K. Kuipers, H. J. Joosten, W. J. Van Berkel, N. G. Leferink, E. Rooijen, E. Ittmann, F. Van Zimmeren, H. Jochens, U. Bornscheuer, G. Vriend, V. A. Martins Dos Santos, and P. J. Schaap, "3dm: Systematic analysis of heterogeneous superfamily data to discover protein functionalities," *Proteins: Structure, Function and Bioinformatics*, vol. 78, no. 9, pp. 2101–2113, 2010.

[40] R. K. P. Kuipers, H. J. Joosten, E. Verwiel, S. Paans, J. Akerboom, J. Van Der Oost, N. G. H. Leferink, W. J. H. Van Berkel, G. Vriend, and P. J. Schaap, "Correlated mutation analyses on super-family alignments reveal functionally important residues," *Proteins: Structure, Function and Bioinformatics*, vol. 76, no. 3, pp. 608–616, 2009.

[41] H. J. Joosten, Y. Han, W. Niu, J. Vervoort, D. Dunaway-Mariano, and P. J. Schaap, "Identification of fungal oxaloacetate hydrolyase within the isocitrate lyase/pep mutase enzyme superfamily using a sequence marker-based method," *Proteins: Structure, Function and Genetics*, vol. 70, no. 1, pp. 157–166, 2008.

[42] H. Jochens and U. T. Bornscheuer, "Natural diversity to guide focused directed evolution," *ChemBioChem*, vol. 11, no. 13, pp. 1861–1866, 2010.

[43] F. Chen, E. A. Gaucher, N. A. Leal, D. Hutter, S. A. Havemann, S. Govindarajan, E. A. Ortlund, and S. A. Benner, "Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and snp detection," *Proceedings of the National Academy of Sciences*, vol. 107, no. 5, pp. 1948–1953, 2010.

[44] R. J. Fox, S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam, S. Muley, J. Grate, J. Gruber, J. C. Whitman, R. A. Sheldon, and G. W. Huisman, "Improving catalytic function by prosar-driven enzyme evolution," *Nature Biotechnology*, vol. 25, no. 3, pp. 338–344, 2007.

[45] Y. Li and P. C. Cirino, "Recent advances in engineering proteins for biocatalysis," *Biotechnology and Bioengineering*, vol. 111, no. 7, pp. 1273–1287, 2014.

[46] N. S. Scrutton, A. Berry, and R. N. Perham, "Redesign of the coenzyme specificity of a dehydrogenase by protein engineering," *Nature*, vol. 343, no. 6253, pp. 38–43, 1990.

[47] S. Hasegawa, K. Uematsu, Y. Natsuma, M. Suda, K. Hiraga, T. Jojima, M. Inui, and H. Yukawa, "Improvement of the redox balance increases l-valine production by corynebacterium glutamicum under oxygen deprivation conditions," *Applied and Environmental Microbiology*, vol. 78, no. 3, pp. 865–875, 2012.

[48] S. Bastian, X. Liu, J. T. Meyerowitz, C. D. Snow, M. M. Chen, and F. H. Arnold, "Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in escherichia coli," *Metabolic Engineering*, vol. 13, no. 3, pp. 345–352, 2011.

[49] S. Wu, J. P. Acevedo, and M. T. Reetz, "Induced allostery in the directed evolution of an enantioselective baeyer-villiger monooxygenase," *Proceedings of the National Academy of Sciences*, vol. 107, no. 7, pp. 2775–2780, 2010.

[50] M. W. Fraaije, J. Wu, D. P. Heuts, E. W. Van Hellemond, J. H. Spelberg, and D. B. Janssen, "Discovery of a thermostable baeyer-villiger monooxygenase by genome mining," *Applied Microbiology and Biotechnology*, vol. 66, no. 4, pp. 393–400, 2005.

[51] C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, Sarena Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman, and G. J. Hughes, "Biocatalytic asymmetric synthesis of sitagliptin manufacture," *Science*, vol. 329, no. July, pp. 305–309, 2010.

[52] M. T. Reetz and H. Zheng, "Manipulating the expression rate and enantioselectivity of an epoxide hydrolase by using directed evolution," *ChemBioChem*, vol. 12, no. 10, pp. 1529–1535, 2011.

[53] S. Kille, F. E. Zilly, J. P. Acevedo, and M. T. Reetz, "Regio- and stereoselectivity of p450-catalysed hydroxylation of steroids controlled by laboratory evolution," *Nature Chemistry*, vol. 3, no. 9, pp. 738–743, 2011.

[54] S. Prasad, M. Bocola, and M. T. Reetz, "Revisiting the lipase from pseudomonas aeruginosa: Directed evolution of substrate acceptance and enantioselectivity using iterative saturation mutagenesis," *ChemPhysChem*, vol. 12, no. 8, pp. 1550–1557, 2011.

[55] Q. Wu, P. Soni, and M. T. Reetz, "Laboratory evolution of enantiocomplementary candida antarctica lipase b mutants with broad substrate scope," *Journal of the American Chemical Society*, vol. 135, no. 5, pp. 1872–1881, 2013.

[56] H. Zheng and M. T. Reetz, "Manipulating the stereoselectivity of limonene epoxide hydrolase by directed evolution based on iterative saturation mutagenesis," *Journal of the American Chemical Society*, vol. 132, no. 44, pp. 15744–15751, 2010.

[57] C. R. Otey, M. Landwehr, J. B. Endelman, K. Hiraga, J. D. Bloom, and F. H. Arnold, "Structure-guided recombination creates an artificial family of cytochromes p450," *PLoS Biology*, vol. 4, no. 5, pp. 789–798, 2006.

[58] Y. Li, D. A. Drummond, A. M. Sawayama, C. D. Snow, J. D. Bloom, and F. H. Arnold, "A diverse family of thermostable cytochrome p450s created by recombination of stabilizing fragments," *Nature Biotechnology*, vol. 25, no. 9, pp. 1051–1056, 2007.

[59] M. M. Meyer, L. Hochrein, and F. H. Arnold, "Structure-guided schema recombination of distantly related β-lactamases," *Protein Engineering, Design and Selection*, vol. 19, no. 12, pp. 563–570, 2006.

[60] P. Heinzelman, C. D. Snow, I. Wu, C. Nguyen, A. Villalobos, S. Govindarajan, J. Minshull, and F. H. Arnold, "A family of thermostable fungal cellulases created by structure-guided recombination," *Proceedings of the National Academy of Sciences*, vol. 106, no. 14, pp. 5610–5615, 2009.

[61] P. Heinzelman, C. D. Snow, M. A. Smith, X. Yu, A. Kannan, K. Boulware, A. Villalobos, S. Govindarajan, J. Minshull, and F. H. Arnold, "Schema recombination of a fungal cellulase uncovers a single mutation that contributes markedly to stability," *Journal of Biological Chemistry*, vol. 284, no. 39, pp. 26229–26233, 2009.

[62] M. Diéguez, J.-E. Bäckvall, and O. Pàmies, *Artificial Metalloenzymes and MetalloDNAzymes in catalysis: from design to applications*. John Wiley & Sons, 2018.

[63] F. Cedrone, A. Ménez, and E. Quéméneur, "Tailoring new enzyme functions by rational redesign," *Current Opinion in Structural Biology*, pp. 405–410, 2000.

[64] B. G. Davis, "Chemical modification of biocatalysts," *Current Opinion in Biotechnology*, vol. 14, no. 4, pp. 379–386, 2003.

[65] K. Zhang, H. Li, K. M. Cho, and J. C. Liao, "Expanding metabolism for total biosynthesis of the nonnatural amino acid l-homoalanine," *Proceedings of the National Academy of Sciences*, vol. 107, no. 14, pp. 6234–6239, 2010.

[66] T. W. Johannes and H. Zhao, "Directed evolution of enzymes and biosynthetic pathways," *Current Opinion in Microbiology*, vol. 9, no. 3, pp. 261–267, 2006.

[67] F. H. Arnold, "How proteins adapt: Lessons from directed evolution," *Trends in Biochemical Sciences*, vol. 74, no. 2, pp. 41–46, 2009.

[68] F. Valetti and G. Gilardi, "Directed evolution of enzymes for product chemistry.," *Natural Product Reports*, vol. 21, no. 4, pp. 490–511, 2004.

[69] M. E. Glasner, J. a. Gerlt, and P. C. Babbitt, "Mechanisms of protein evolution and their application to protein engineering.," *Advances in Enzymology and Related Areas of Molecular Biology*, vol. 75, pp. 193—-239, xii–xiii, 2007.

[70] S. Lutz, "Beyond directed evolution-semi-rational protein engineering and design," *Current Opinion in Biotechnology*, vol. 21, no. 6, pp. 734–743, 2010.

[71] P. Banáš, M. Otyepka, P. Jeřábek, M. Petřek, and J. Damborsky, "Mechanism of enhanced conversion of 1,2,3-trichloropropane by mutant haloalkane dehalogenase revealed by molecular modeling," *Journal of Computer-Aided Molecular Design*, vol. 20, no. 6, pp. 375–383, 2006.

[72] E. Malito, A. Alfieri, M. W. Fraaije, and A. Mattevi, "Crystal structure of a baeyer-villiger monooxygenase.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 36, pp. 13157–62, 2004.

[73] K. Wieszczycka and K. Staszak, "Artificial metalloenzymes as catalysts in non-natural compounds synthesis," *Coordination Chemistry Reviews*, vol. 351, pp. 160–171, 2017.

[74] I. Bertini, H. B. Gray, E. I. Stiefel, and J. S. Valentine, "Biological inorganic chemistry: structure and reactivity," *Angewandte Chemie - International Edition*, vol. 46, no. 46, pp. 8741–8742, 2006.

[75] A. F. Peacock, "Incorporating metals into de novo proteins," *Current Opinion in Chemical Biology*, vol. 17, no. 6, pp. 934–939, 2013.

[76] M. Hoppert, "Metalloenzymes," in *Encyclopedia of Geobiology*, pp. 558–563, Springer, 2011.

[77] B. J. Herring, A. L. Logsdon, J. E. Lockard, B. M. Miller, H. Kim, E. a. Calderon, J. Vincent, and M. M. Bailey, "Long-term exposure to $[Cr_3O(O_2CCH_2CH_3)_6(H_2O)_3)]^+$ in wistar rats fed normal or high-fat diets does not alter glucose metabolism.," *Biological Trace Element Research*, vol. 151, no. 3, pp. 406–14, 2013.

[78] A. G. Tebo and V. L. Pecoraro, "Artificial metalloenzymes derived from three-helix bundles," *Current Opinion in Chemical Biology*, vol. 25, pp. 65–70, 2015.

[79] I. Drienovská, A. Rioz-Martínez, A. Draksharapu, and G. Roelfes, "Novel artificial metalloenzymes by *in vivo* incorporation of metal-binding unnatural amino acids," *Chemical Science*, vol. 6, no. 1, pp. 770–776, 2015.

[80] A. Pordea, "Metal-binding promiscuity in artificial metalloenzyme design," *Current Opinion in Chemical Biology*, vol. 25, pp. 124–132, 2015.

[81] J. Liu, K. K. Meier, S. Tian, J. L. Zhang, H. Guo, C. E. Schulz, H. Robinson, M. J. Nilges, E. Münck, and Y. Lu, "Redesigning the blue copper azurin into a redox-active mononuclear nonheme iron protein: Preparation and study of fe(ii)-m121e azurin," *Journal of the American Chemical Society*, vol. 136, no. 35, pp. 12337–12344, 2014.

[82] X. Liu, Y. Yu, C. Hu, W. Zhang, Y. Lu, and J. Wang, "Significant increase of oxidase activity through the genetic incorporation of a tyrosine-histidine cross-link in a myoglobin model of heme-copper oxidase," *Angewandte Chemie - International Edition*, vol. 51, no. 18, pp. 4312–4316, 2012.

[83] M. Hoarau, C. Hureau, E. Gras, and P. Faller, "Coordination complexes and biomolecules: A wise wedding for catalysis upgrade," *Coordination Chemistry Reviews*, vol. 308, pp. 445–459, 2016.

[84] F. H. Arnold, "The nature of chemical innovation: New enzymes by evolution," *Quarterly Reviews of Biophysics*, vol. 48, no. 4, pp. 404–410, 2015.

[85] C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday, and J. M. Thornton, "Metal ions in biological catalysis: From enzyme databases to general principles," *Journal of Biological Inorganic Chemistry*, vol. 13, no. 8, pp. 1205–1218, 2008.

[86] V. Muñoz Robles, E. Ortega-Carrasco, L. Alonso-Cotchico, J. Rodriguez-Guerra, A. Lledós, and J.-D. Maréchal, "Toward the computational design of artificial metalloenzymes: From protein-ligand docking to multiscale approaches," *ACS Catalysis*, vol. 5, no. 4, pp. 2469–2480, 2015.

[87] F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Köhler, J. C. Lewis, and T. R. Ward, "Artificial metalloenzymes: Reaction scope and optimization strategies," *Chemical Reviews*, p. acs.chemrev.7b00014, 2017.

[88] T. Heinisch and T. R. Ward, "Artificial metalloenzymes based on the biotin-streptavidin technology: Challenges and opportunities," *Accounts of Chemical Research*, vol. 49, no. 9, pp. 1711–1721, 2016.

[89] A. Mahammed and Z. Gross, "Albumin-conjugated corrole metal complexes: Extremely simple yet very efficient biomimetic oxidation systems," *Journal of the American Chemical Society*, vol. 127, no. 9, pp. 2883–2887, 2005.

[90] M. T. Reetz and N. Jiao, "Copper-phthalocyanine conjugates of serum albumins as enantioselective catalysts in diels-alder reactions," *Angewandte Chemie - International Edition*, vol. 45, no. 15, pp. 2416–2419, 2006.

[91] T. Kokubo, T. Sugimoto, T. Uchida, S. Tanimoto, and M. Okano, "The bovine serum albumin–2-phenylpropane-1,2-diolatodioxo-osmium(vi) complex as an enantioselective catalyst for cis-hydroxylation of alkenes," *Chemical Communications*, no. 14, pp. 769–770, 1983.

[92] E. W. Reynolds, T. D. Schwochert, M. W. McHenry, J. W. Watters, and E. M. Brustad, "Orthogonal expression of an artificial metalloenzyme for abiotic catalysis," *ChemBioChem*, vol. 18, no. 24, pp. 2380–2384, 2017.

[93] W. Ghattas, J.-P. Mahy, R. Ricoux, J.-D. Maréchal, A. Urvoas, L. Cotchico-Alonso, and M. Rousseau, "Artificial metalloenzymes with the neocarzinostatin-scaffold : toward a biocatalyst for the diels-alder cyclization reaction.," *ChemBioChem*, vol. 17, no. 5, pp. 433–440, 2015.

[94] J. Bos, A. García-Herraiz, and G. Roelfes, "An enantioselective artificial metallo-hydratase," *Chemical Science*, vol. 4, no. 9, p. 3578, 2013.

[95] I. Drienovská, L. Alonso-Cotchico, P. Vidossich, A. Lledós, J.-D. Maréchal, and G. Roelfes, "Design of an enantioselective artificial metallo-hydratase enzyme containing an unnatural metal-binding amino acid," *Chemical Science*, vol. 8, no. 10, pp. 7228–7235, 2017.

[96] A. Rioz-Martínez and G. Roelfes, "Dna-based hybrid catalysis," *Current Opinion in Chemical Biology*, vol. 25, pp. 80–87, 2015.

[97] S. Dey, C. L. Rühl, and A. Jäschke, "Catalysis of michael additions by covalently modified g-quadruplex dna," *Chemistry*, vol. 23, no. 50, pp. 12162–12170, 2017.

[98] A. García-Fernández, R. P. Megens, L. Villarino, and G. Roelfes, "Dna-accelerated copper catalysis of friedel-crafts conjugate addition/enantioselective protonation reactions in water," *Journal of the American Chemical Society*, vol. 138, no. 50, pp. 16308–16314, 2016.

[99] A. J. Boersma, B. L. Feringa, and G. Roelfes, "Enantioselective friedel-crafts reactions in water using a dna*based catalyst," *Angewandte Chemie - International Edition*, vol. 48, no. 18, pp. 3346–3348, 2009.

[100] A. Rioz-Martínez, J. Oelerich, N. Ségaud, and G. Roelfes, "Dna-accelerated catalysis of carbene-transfer reactions by a dna/cationic iron porphyrin hybrid," *Angewandte Chemie - International Edition*, vol. 55, no. 45, pp. 14136–14140, 2016.

[101] M. Ohashi, T. Koshiyama, T. Ueno, M. Yanase, H. Fujii, and Y. Watanabe, "Preparation of artificial metalloenzymes by insertion of chromium(iii) schiff base complexes into apomyoglobin mutants," *Angewandte Chemie - International Edition*, vol. 42, no. 9, pp. 1005–1008, 2003.

[102] M. T. Reetz, M. Rentzsch, A. Pletsch, and M. Maywald, "Towards the directed evolution of hybrid catalysts," *CHIMIA International Journal for Chemistry*, vol. 56, no. 12, pp. 721–723, 2002.

[103] F. Van De Velde, I. W. Arends, and R. A. Sheldon, "Biocatalytic and biomimetic oxidations with vanadium," *Journal of Inorganic Biochemistry*, vol. 80, no. 1-2, pp. 81–89, 2000.

[104] A. Chatterjee, H. Mallin, J. Klehr, J. Vallapurackal, A. D. Finke, L. Vera, M. Marsh, and T. R. Ward, "An enantioselective artificial suzukiase based on

the biotin–streptavidin technology," *Chemical Science*, vol. 7, no. 1, pp. 673–677, 2016.

[105] M. Hestericová, T. Heinisch, L. Alonso-Cotchico, J. D. Maréchal, P. Vidossich, and T. R. Ward, "Directed evolution of an artificial imine reductase," *Angewandte Chemie - International Edition*, vol. 57, no. 7, pp. 1863–1868, 2018.

[106] W. Ghattas, L. Cotchico-Alonso, J. D. Maréchal, A. Urvoas, M. Rousseau, J. P. Mahy, and R. Ricoux, "Artificial metalloenzymes with the neocarzinostatin scaffold: Toward a biocatalyst for the diels-alder reaction," *ChemBioChem*, vol. 17, no. 5, pp. 433–440, 2016.

[107] Q. Jing, K. Okrasa, and R. J. Kazlauskas, "Manganese-substituted α-carbonic anhydrase as an enantioselective peroxidase," *Topics in Organometallic Chemistry*, vol. 25, pp. 45–61, 2009.

[108] S. Akabori, S. Sakurai, Y. Izumi, and Y. Fujii, "An asymmetric catalyst," *Nature*, vol. 178, no. 4528, pp. 323–324, 1956.

[109] J. E. Coleman, "Metal ion dependent binding of sulphonamide to carbonic anhydrase," *Nature*, vol. 214, no. 5084, pp. 193–194, 1967.

[110] K. Yamamura and E. T. Kaiser, "Studies on the oxidase activity of copper(ii) carboxypeptidase a," *Chemical Communications*, no. 20, p. 830, 1976.

[111] P. Cuatrecasas, S. Fuchs, and C. B. Anfinsen, "Catalytic properties and specificity of the extracellular nuclease of staphylococcus aureus.," *Journal of Biological Chemistry*, vol. 242, no. 7, pp. 1541–1547, 1967.

[112] M. E. Wilson and G. M. Whitesides, "Conversion of a protein to a homogeneous asymmetric hydrogenation catalyst by site-specific modification with a diphos-

phinerhodium(i) moiety," *Journal of the American Chemical Society*, vol. 100, pp. 306–307, 1978.

[113] T. Sugimoto, T. Kokubo, J. Miyazaki, S. Tanimoto, and M. Okano, "Stereoselective oxidation of aromatic sulfides and sulfoxides in the binding domain of bovine serum albumin," *Bioorganic Chemistry*, vol. 10, no. 3, pp. 311–323, 1981.

[114] J. C. Lewis, "Artificial metalloenzymes and metallopeptide catalysts for organic synthesis," *ACS Catalysis*, vol. 3, no. 12, pp. 2954–2975, 2013.

[115] M. Jeschek, S. Panke, and T. R. Ward, "Artificial metalloenzymes on the verge of new-to-nature metabolism," *Trends in Biotechnology*, vol. 36, no. 1, pp. 60–72, 2018.

[116] G. Kiss, N. Çelebi-Ölçüm, R. Moretti, D. Baker, and K. Houk, "Computational enzyme design," *Angewandte Chemie - International Edition*, vol. 52, no. 22, pp. 5700–5725, 2013.

[117] D. N. Bolon and S. L. Mayo, "Enzyme-like proteins by computational design," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 25, pp. 14274–14279, 2001.

[118] D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, and D. Baker, "Kemp elimination catalysts by computational enzyme design.," *Nature*, vol. 453, no. 7192, pp. 190–5, 2008.

[119] J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael, and D. Baker, "Computational design of an enzyme catalyst for

a stereoselective bimolecular diels-alder reaction.," *Science*, vol. 329, no. 5989, pp. 309–13, 2010.

[120] L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Rothlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas, D. Hilvert, K. N. Houk, B. L. Stoddard, and D. Baker, "De novo computational design of retro-aldol enzymes," *Science*, vol. 319, no. 5868, pp. 1387–1391, 2008.

[121] A. Zanghellini, L. Jiang, A. M. Wollacott, G. Cheng, J. Meiler, E. A. Althoff, D. Röthlisberger, and D. Baker, "New algorithms and an in silico benchmark for computational enzyme design.," *Protein Science*, vol. 15, no. 12, pp. 2785–2794, 2006.

[122] O. Khersonsky, D. Röthlisberger, O. Dym, S. Albeck, C. J. Jackson, D. Baker, and D. S. Tawfik, "Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the ke07 series," *Journal of Molecular Biology*, vol. 396, no. 4, pp. 1025–1042, 2010.

[123] O. Khersonsky, D. Röthlisberger, A. M. Wollacott, P. Murphy, O. Dym, S. Albeck, G. Kiss, K. N. Houk, D. Baker, and D. S. Tawfik, "Optimization of the in-silico-designed kemp eliminase ke70 by computational design and directed evolution.," *Journal of Molecular Biology*, vol. 407, no. 3, pp. 391–412, 2011.

[124] O. Khersonsky, G. Kiss, D. Rothlisberger, O. Dym, S. Albeck, K. N. Houk, D. Baker, and D. S. Tawfik, "Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed kemp eliminase ke59," *Proceedings of the National Academy of Sciences*, vol. 109, no. 26, pp. 10358–10363, 2012.

[125] H. K. Privett, G. Kiss, T. M. Lee, R. Blomberg, R. A. Chica, L. M. Thomas, D. Hilvert, K. N. Houk, and S. L. Mayo, "Iterative approach to computational

enzyme design.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 10, pp. 3790–3795, 2012.

[126] S. Bjelic, L. G. Nivón, N. Çelebi-Ölçüm, G. Kiss, C. F. Rosewall, H. M. Lovick, E. L. Ingalls, J. L. Gallaher, J. Seetharaman, S. Lew, G. T. Montelione, J. F. Hunt, F. E. Michael, K. N. Houk, and D. Baker, "Computational design of enone-binding proteins with catalytic activity for the morita-baylis-hillman reaction," *ACS Chemical Biology*, vol. 8, no. 4, pp. 749–757, 2013.

[127] F. Richter, A. Leaver-Fay, S. D. Khare, S. Bjelic, and D. Baker, "De novo enzyme design using rosetta3.," *PloS one*, vol. 6, no. 5, p. doi: 10.1371/journal.pone.0019230, 2011.

[128] F. Richter, R. Blomberg, S. D. Khare, G. Kiss, A. P. Kuzin, A. J. Smith, J. Gallaher, Z. Pianowski, R. C. Helgeson, A. Grjasnow, R. Xiao, J. Seetharaman, M. Su, S. Vorobiev, S. Lew, F. Forouhar, G. J. Kornhaber, J. F. Hunt, G. T. Montelione, L. Tong, K. N. Houk, D. Hilvert, and D. Baker, "Computational design of catalytic dyads and oxyanion holes for ester hydrolysis," *Journal of the American Chemical Society*, vol. 134, no. 39, pp. 16197–16206, 2012.

[129] J. Z. Ruscio, J. E. Kohn, K. A. Ball, and T. Head-Gordon, "The influence of protein dynamics on the success of computational enzyme design," *Journal of the American Chemical Society*, vol. 131, no. 39, pp. 14111–14115, 2009.

[130] M. Schneider, X. Fu, and A. E. Keating, "X-ray vs. nmr structures as templates for computational protein design," *Proteins: Structure, Function and Bioinformatics*, vol. 77, no. 1, pp. 97–110, 2009.

[131] N. Yeung, Y.-W. Lin, Y.-G. Gao, X. Zhao, B. S. Russell, L. Lei, K. D. Miner, H. Robinson, and Y. Lu, "Rational design of a structural and functional nitric oxide reductase," *Nature*, vol. 462, no. 7276, pp. 1079–1082, 2009.

[132] Y.-W. Lin, N. Yeung, Y.-G. Gao, K. D. Miner, S. Tian, H. Robinson, and Y. Lu, "Roles of glutamates and metal ions in a rationally designed nitric oxide reductase based on myoglobin," *Proceedings of the National Academy of Sciences*, vol. 107, no. 19, pp. 8581–8586, 2010.

[133] D. J. Tantillo, J. Chen, and K. N. Houk, "Theozymes and compuzymes: theoretical models for biological catalysis.," *Current Opinion in Chemical Biology*, vol. 2, no. 6, pp. 743–50, 1998.

[134] H. M. Berman, "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.

[135] H. W. Hellinga and F. M. Richards, "Construction of new ligand binding sites in proteins of known structure. i. computer-aided modeling of sites with predefined geometry.," *Journal of Molecular Biology*, vol. 222, no. 3, pp. 763–85, 1991.

[136] B. I. Dahiyat and S. L. Mayo, "Protein design automation.," *Protein Science*, vol. 5, no. 5, pp. 895–903, 1996.

[137] H. Fazelinia, P. C. Cirino, and C. D. Maranas, "Optgraft: A computational procedure for transferring a binding site onto an existing protein scaffold.," *Protein Science*, vol. 18, no. 1, pp. 180–95, 2009.

[138] Y. Lei, W. Luo, and Y. Zhu, "A matching algorithm for catalytic residue site selection in computational enzyme design.," *Protein Science*, vol. 20, no. 9, pp. 1566–75, 2011.

[139] G. R. Nosrati and K. N. Houk, "Saber: a computational method for identifying active sites for new reactions.," *Protein Science*, vol. 21, no. 5, pp. 697–706, 2012.

[140] C. Malisi, O. Kohlbacher, and B. Höcker, "Automated scaffold selection for enzyme design.," *Proteins*, vol. 77, no. 1, pp. 74–83, 2009.

[141] S. Hu, J. Huang, L. Mei, Q. Yu, S. Yao, and Z. Jin, "Altering the regioselectivity of cytochrome p450 bm-3 by saturation mutagenesis for the biosynthesis of indirubin," *Journal of Molecular Catalysis B: Enzymatic*, vol. 67, no. 1-2, pp. 29–35, 2010.

[142] S. M. Lippow, T. S. Moon, S. Basu, S. H. Yoon, X. Li, B. A. Chapman, K. Robison, D. Lipovsek, and K. L. J. Prather, "Engineering enzyme specificity using computational design of a defined-sequence library," *Chemistry and Biology*, vol. 17, no. 12, pp. 1306–1315, 2010.

[143] M. Dippe, W. Brandt, H. Rost, A. Porzel, J. Schmidt, and L. A. Wessjohann, "Rationally engineered variants of s-adenosylmethionine (sam) synthase: reduced product inhibition and synthesis of artificial cofactor homologues," *Chemical Communications*, vol. 51, no. 17, pp. 3637–3640, 2015.

[144] J. Damborsky and J. Brezovsky, "Computational tools for designing and engineering enzymes," *Current Opinion in Chemical Biology*, vol. 19, no. 1, pp. 8–16, 2014.

[145] J. Nyhlén, B. Martín-Matute, A. G. Sandström, M. Bocola, and J.-E. Bäckvall, "Influence of delta-functional groups on the enantiorecognition of secondary alcohols by candida antarctica lipase b.," *ChemBiochem*, vol. 9, no. 12, pp. 1968–74, 2008.

[146] P. Braiuca, C. Ebert, A. Basso, P. Linda, and L. Gardossi, "Computational methods to rationalize experimental strategies in biocatalysis," *Trends in Biotechnology*, vol. 24, no. 9, pp. 419–425, 2006.

[147] M. T. Reetz, M. Puls, J. D. Carballeira, A. Vogel, K. E. Jaeger, T. Eggert, W. Thiel, M. Bocola, and N. Otte, "Learning from directed evolution: Further lessons from theoretical investigations into cooperative mutations in lipase enantioselectivity," *ChemBioChem*, vol. 8, no. 1, pp. 106–112, 2007.

[148] Z. Ni, P. Zhou, X. Jin, and X. F. Lin, "Integrating in silico and in vitro approaches to dissect the stereoselectivity of bacillus subtilis lipase a toward ketoprofen vinyl ester," *Chemical Biology and Drug Design*, vol. 78, no. 2, pp. 301–308, 2011.

[149] M. Mladenovic, K. Ansorg, R. F. Fink, W. Thiel, T. Schirmeister, and B. Engels, "Atomistic insights into the inhibition of cysteine proteases: First qm/mm calculations clarifying the stereoselectivity of epoxide-based inhibitors," *Journal of Physical Chemistry B*, vol. 112, no. 37, pp. 11798–11808, 2008.

[150] D. Li, X. Huang, K. Han, and C. G. Zhan, "Catalytic mechanism of cytochrome p450 for 5???-hydroxylation of nicotine: Fundamental reaction pathways and stereoselectivity," *Journal of the American Chemical Society*, vol. 133, no. 19, pp. 7416–7427, 2011.

[151] S. Osuna, G. Jiménez-Osés, E. L. Noey, and K. N. Houk, "Molecular dynamics explorations of active site structure in designed and evolved enzymes," *Accounts of Chemical Research*, 2015.

[152] S. Acebes, E. Fernandez-Fueyo, E. Monza, M. F. Lucas, D. Almendral, F. J. Ruiz-Dueñas, H. Lund, A. T. Martinez, and V. Guallar, "Rational enzyme engineering through biophysical and biochemical modeling," *ACS Catalysis*, pp. 1624–1629, 2016.

[153] E. L. Noey, N. Tibrewal, G. Jiménez-Osés, S. Osuna, J. Park, C. M. Bond, D. Cascio, J. Liang, X. Zhang, G. W. Huisman, Y. Tang, and K. N. Houk,

"Origins of stereoselectivity in evolved ketoreductases," *Proceedings of the National Academy of Sciences*, p. 201507910, 2015.

[154] H. J. Wijma, R. J. Floor, P. A. Jekel, D. Baker, S. J. Marrink, and D. B. Janssen, "Computationally designed libraries for rapid enzyme stabilization," *Protein Engineering, Design and Selection*, vol. 27, no. 2, pp. 49–58, 2014.

[155] M. J. Grisewood, N. P. Gifford, R. J. Pantazes, Y. Li, P. C. Cirino, M. J. Janik, and C. D. Maranas, "Optzyme: Computational enzyme redesign using transition state analogues," *PLoS ONE*, vol. 8, no. 10, 2013.

[156] C.-Y. Chen, I. Georgiev, A. C. Anderson, and B. R. Donald, "Computational structure-based redesign of enzyme activity," *Proceedings of the National Academy of Sciences*, vol. 106, no. 10, pp. 3764–3769, 2009.

[157] R. Das and D. Baker, "Macromolecular modeling with rosetta," *Annual Review of Biochemistry*, vol. 77, no. 1, pp. 363–382, 2008.

[158] P. M. Murphy, J. M. Bolduc, J. L. Gallaher, B. L. Stoddard, and D. Baker, "Alteration of enzyme specificity by computational loop remodeling and design," *Proceedings of the National Academy of Sciences*, vol. 106, no. 23, pp. 9215–9220, 2009.

[159] X. Liu, X. Wang, and H. Jiang, "A steered molecular dynamics method with direction optimization and its applications on ligand molecule dissociation.," *Journal of Biochemical and Biophysical Methods*, vol. 70, no. 6, pp. 857–864, 2008.

[160] A. N. Lima, E. A. Philot, D. Perahia, A. S. K. Braz, and L. P. B. Scott, "Ganm: A protein-ligand docking approach based on genetic algorithm and normal modes," *Applied Mathematics and Computation*, vol. 219, no. 2, pp. 511–520, 2012.

[161] K. W. Borrelli, A. Vitalis, R. Alcantara, and V. Guallar, "Pele: Protein energy landscape exploration. a novel monte carlo based technique," *Journal of Chemical Theory and Computation*, vol. 1, no. 6, pp. 1304–1311, 2005.

[162] C. Cavazza, C. Bochot, P. Rousselot-Pailley, P. Carpentier, M. V. Cherrier, L. Martin, C. Marchi-Delapierre, J. C. Fontecilla-Camps, and S. Ménage, "Crystallographic snapshots of the reaction of aromatic c-h with o(2) catalysed by a protein-bound iron complex.," *Nature Chemistry*, vol. 2, no. 12, pp. 1069–76, 2010.

[163] C. Esmieu, M. V. Cherrier, P. Amara, E. Girgenti, C. Marchi-Delapierre, F. Oddon, M. Iannello, A. Jorge-Robin, C. Cavazza, and S. Ménage, "An artificial oxygenase built from scratch: substrate binding site identified using a docking approach.," *Angewandte Chemie - International Edition*, vol. 52, no. 14, pp. 3922–3925, 2013.

[164] Z. Ke, S. Abe, T. Ueno, and K. Morokuma, "Rh-catalyzed polymerization of phenylacetylene: Theoretical studies of the reaction mechanism, regioselectivity, and stereoregularity," *Journal of the American Chemical Society*, vol. 133, no. 20, pp. 7926–7941, 2011.

[165] V. Muñoz Robles, E. Ortega-Carrasco, E. G. Fuentes, A. Lledós, and J.-D. Maréchal, "What can molecular modelling bring to the design of artificial inorganic cofactors?," *Faraday Discussions*, vol. 148, p. 137, 2011.

[166] V. Muñoz Robles, P. Vidossich, A. Lledós, T. R. Ward, and J.-D. Maréchal, "Computational insights on an artificial imine reductase based on the biotin-streptavidin technology," *ACS Catalysis*, vol. 4, pp. 833–842, 2014.

[167] M. Allard, C. Dupont, V. Muñoz Robles, N. Doucet, A. Lledós, J. D. Maréchal, A. Urvoas, J. P. Mahy, and R. Ricoux, "Incorporation of manganese complexes

into xylanase: New artificial metalloenzymes for enantioselective epoxidation," *ChemBioChem*, vol. 13, no. 2, pp. 240–251, 2012.

[168] E. Ortega-Carrasco, F. P. Cossío, A. Lledós, and J. D. Maréchal, "Computational insights on the possibility of tri-coordinated cisplatinated adducts with protein models," *Journal of Inorganic Biochemistry*, vol. 117, pp. 230–236, 2012.

[169] E. Ortega-Carrasco, A. Lledos, and J.-D. Marechal, "Unravelling novel synergies between organometallic and biological partners: a quantum mechanics/molecular mechanics study of an artificial metalloenzyme," *Journal of The Royal Society Interface*, vol. 11, no. 96, p. 10.1098/rsif.2014.0090, 2014.

[170] A. de Cózar, O. Larrañaga, F. M. Bickelhaupt, E. San Sebastián, E. Ortega-Carrasco, J. D. Maréchal, A. Lledós, and F. P. Cossío, "New insights into the reactivity of cisplatin with free and restrained nucleophiles: Microsolvation effects and base selectivity in cisplatin–dna interactions," *ChemPhysChem*, vol. 17, no. 23, pp. 3932–3947, 2016.

[171] L. Villarino, K. Splan, E. Reddem, C. Gutiérrez de Souza, L. Alonso-Cotchico, A. Lledós, J.-D. Maréchal, A.-M. Thunnissen, and G. Roelfes, "An artificial heme enzyme for cyclopropanation reactions," *Angewandte Chemie - International Edition*, p. doi: 10.1002/anie.201802946, 2018.

[172] J.-P. Mahy, J.-D. Maréchal, and R. Ricoux, "From "hemoabzymes" to "hemozymes": towards new biocatalysts for selective oxidations," *Chemical Communications*, vol. 51, no. 13, pp. 2476–2494, 2015.

[173] D. C. Young, *Spin Contamination*. John Wiley & Sons, 2001.

[174] A. Leach, "Molecular modelling: principles and applications," *Computers*, vol. 21, no. 3, p. 784, 2001.

[175] C. J. Cramer, *Essentials of Computational Chemistry Theories and Models*, vol. 42. John Wiley & Sons, 2004.

[176] F. Jensen, *Introduction to computational chemistry*. John Wiley & Wons, 2017.

[177] I. Tuñón and V. Moliner, *Simulating enzyme reactivity: computational methods in enzyme catalysis*. Royal Society of Chemistry, 2016.

[178] W. Koch and M. C. Holthausen, *A chemist's guide to density functional theory*. John Wiley & Sons, 2015.

[179] P. Hohenberg and W. Kohn, "Inhomogeneous eelectron gas," *Physical Review*, vol. 136, no. 3B, pp. B864–B871, 1964.

[180] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Physical Review*, vol. 140, no. 4A, 1965.

[181] J. P. Perdew, "Jacob's ladder of density functional approximations for the exchange-correlation energy," *AIP Conference Proceedings*, vol. 577, pp. 1–20, 2001.

[182] P. Rushton, S. Clark, and D. Tozer, "Density-functional calculations of semiconductor properties using a semiempirical exchange-correlation functional," *Physical Review B*, vol. 63, pp. 1–5, 2001.

[183] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Physical Review A*, vol. 38, no. 6, pp. 3098–3100, 1988.

[184] C. Lee, W. Yang, and R. Parr, "Development of the colle- salvetti correlation energy formula into a functional of the electron density," *Physical Review B*, vol. 37, no. 2, pp. 785–789, 1988.

[185] B. Miehlich, A. Savin, H. Stoll, and H. Preuss, "Results obtained with the correlation energy density functionals of becke and lee, yang and parr," *Chemical Physics Letters*, vol. 157, no. 3, pp. 200–206, 1989.

[186] N. C. Handy and A. J. Cohen, "Left-right correlation energy," *Molecular Physics*, vol. 99, no. 5, pp. 403–412, 2001.

[187] W. M. Hoe, A. J. Cohen, and N. C. Handy, "Assessment of a new local exchange functional optx," *Chemical Physics Letters*, vol. 341, no. 3-4, pp. 319–328, 2001.

[188] M. P. Allen and D. J. Tildesley, "Computer simulation of," *Liquids, Oxford University Press, New York*, vol. 18, no. 195, p. 385, 1987.

[189] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple- errata," *Physical Review Letters*, vol. 77, no. 18, pp. 3865–3868, 1996.

[190] J. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple.," *Physical Review Letters*, vol. 77, no. 18, pp. 3865–3868, 1996.

[191] J. P. Perdew, "Density-functional approximation for the correlation energy of the inhomogeneous electron gas," *Physical Review B*, vol. 33, no. 12, pp. 8822–8824, 1986.

[192] J. P. Perdew, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, "Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation," *Physical Review B*, vol. 46, no. 11, pp. 6671–6687, 1992.

[193] J. Perdew, J. Chevary, S. Vosko, K. Jackson, M. Pederson, D. Singh, and C. Fiolhais, "Erratum: Atoms, molecules, solids, and surfaces: Applications of the

generalized gradient approximation for exchange and correlation," *Physical Review B*, vol. 48, no. 7, pp. 4978–4978, 1993.

[194] J. P. Perdew and K. Burke, "Generalized gradient approximation for the exchange-correlation hole of a many-electron system," *Physical Review B*, vol. 54, no. 23, pp. 16533–16539, 1996.

[195] A. D. Becke and M. R. Roussel, "Exchange holes in inhomogeneous systems: A coordinate-space model," *Physical Review A*, vol. 39, no. 8, pp. 3761–3767, 1989.

[196] A. D. Becke, "Density-functional thermochemistry. iv. a new dynamical correlation functional and implications for exact-exchange mixing," *Journal of Chemical Physics*, vol. 104, no. 3, pp. 1040–1046, 1996.

[197] A. D. Boese and N. C. Handy, "New exchange-correlation density functionals: The role of the kinetic-energy density," *Journal of Chemical Physics*, vol. 116, no. 22, pp. 9559–9569, 2002.

[198] T. Van Voorhis and G. E. Scuseria, "A novel form for the exchange-correlation energy functional," *Journal of Chemical Physics*, vol. 109, no. 2, pp. 400–410, 1998.

[199] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, "Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids," *Physical Review Letters*, vol. 91, no. 14, 2003.

[200] A. D. Becke, "Density-functional thermochemistry. iii. the role of exact exchange," *Journal of Chemical Physics*, vol. 98, no. 7, pp. 5648–5652, 1993.

[201] 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," *Journal of Physical Chemistry*, vol. 98, no. 45, pp. 11623–11627, 1994.

[202] S. Grimme, J. Antony, S. Ehrlich, and H. 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*, vol. 132, no. 15, 2010.

[203] H. L. Schmider and A. D. Becke, "Optimized density functionals from the extended g2 test set," *Journal of Chemical Physics*, vol. 108, no. 23, pp. 9624–9631, 1998.

[204] M. Ernzerhof and G. E. Scuseria, "Assessment of the perdew-burke-ernzerhof exchange-correlation functional," *Journal of Chemical Physics*, vol. 110, no. 11, pp. 5029–5036, 1999.

[205] V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, "Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes," *Journal of Chemical Physics*, vol. 119, no. 23, pp. 12129–12137, 2003.

[206] V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, "Erratum: "comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes" [j. chem. phys. 119, 12129 (2003)]," *Journal of Chemical Physics*, vol. 121, no. 22, p. 11507, 2004.

[207] A. K. Rappé and C. J. Casewit, *Molecular mechanics across chemistry*. University Science Books, 1997.

[208] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A sec-

ond generation force field for the simulation of proteins, nucleic acids, and organic molecules," *Journal of the American Chemical Society*, vol. 117, no. 19, pp. 5179–5197, 1995.

[209] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "Charmm: A program for macromolecular energy, minimization, and dynamics calculations," *Journal of Computational Chemistry*, vol. 4, no. 2, pp. 187–217, 1983.

[210] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *Journal of Physical Chemistry B*, vol. 102, no. 18, pp. 3586–3616, 1998.

[211] R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig, and A. D. MacKerell, "Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ1and χ2dihedral angles," *Journal of Chemical Theory and Computation*, vol. 8, no. 9, pp. 3257–3273, 2012.

[212] A. D. MacKerell, J. Wiórkiewicz-Kuczera, M. Karplus, and A. D. MacKerell, "An all-atom empirical energy function for the simulation of nucleic acids," *Journal of the American Chemical Society*, vol. 117, no. 48, pp. 11946–11975, 1995.

[213] N. Foloppe and A. D. MacKerell, "All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data," *Journal of Computational Chemistry*, vol. 21, no. 2, pp. 86–104, 2000.

[214] A. D. MacKerell and N. K. Banavali, "All-atom empirical force field for nucleic acids: Ii. application to molecular dynamics simulations of dna and rna in solution," *Journal of Computational Chemistry*, vol. 21, no. 2, pp. 105–120, 2000.

[215] S. E. Feller, A. D. Mackerell Jr, and A. D. J. Mackerell, "An improved empirical potential energy function for molecular simulations of phospholipids supporting 2," *Journal of Physical Chemistry B*, vol. 104, no. 31, pp. 7510–7515, 2000.

[216] S. E. Feller, K. Gawrisch, and A. D. MacKerell, "Polyunsaturated fatty acids in lipid bilayers: Intrinsic and environmental contributions to their unique physical properties," *Journal of the American Chemical Society*, vol. 124, no. 2, pp. 318–326, 2002.

[217] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell, and R. W. Pastor, "Update of the charmm all-atom additive force field for lipids: Validation on six lipid types," *Journal of Physical Chemistry B*, vol. 114, no. 23, pp. 7830–7843, 2010.

[218] O. Guvench, S. N. Greenr, G. Kamath, J. W. Brady, R. M. Venable, R. W. Pastor, and A. D. Mackerell, "Additive empirical force field for hexopyranose monosaccharides," *Journal of Computational Chemistry*, vol. 29, no. 15, pp. 2543–2564, 2008.

[219] E. R. Hatcher, O. Guvench, and A. D. MacKerell, "Charmm additive all-atom force field for acyclic polyalcohols, acyclic carbohydrates, and inositol," *Journal of Chemical Theory and Computation*, vol. 5, no. 5, pp. 1315–1327, 2009.

[220] W. R. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, and W. F. Van Gunsteren, "The gromos biomolecular simulation program package," *Journal of Physical Chemistry A*, vol. 103, no. 19, pp. 3596–3607, 1999.

[221] P. E. M. Lopes, O. Guvench, and A. D. Mackerell, "Current status of protein force fields for molecular dynamics simulations," *Molecular Modelling of Proteins*, vol. 1215, pp. 47–71, 2015.

[222] 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*, vol. 31, no. 4, pp. 671–690, 2010.

[223] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *Journal of Computational Chemistry*, vol. 25, no. 9, pp. 1157–1174, 2004.

[224] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple amber force fields and development of improved protein backbone parameters," *Proteins: Structure, Function and Genetics*, vol. 65, no. 3, pp. 712–725, 2006.

[225] R. B. Best and G. Hummer, "Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides," *Journal of Physical Chemistry B*, vol. 113, no. 26, pp. 9004–9015, 2009.

[226] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the amber ff99sb protein force field," *Proteins: Structure, Function and Bioinformatics*, vol. 78, no. 8, pp. 1950–1958, 2010.

[227] D. W. Li and R. Brüschweiler, "Nmr-based protein potentials," *Angewandte Chemie - International Edition*, vol. 49, no. 38, pp. 6778–6780, 2010.

[228] P. S. Nerenberg and T. Head-Gordon, "Optimizing protein-solvent force fields to reproduce intrinsic conformational preferences of model peptides," *Journal of Chemical Theory and Computation*, vol. 7, no. 4, pp. 1220–1230, 2011.

[229] P. Li and K. M. Merz, "Metal ion modeling using classical mechanics," *Chemical Reviews*, vol. 117, no. 3, pp. 1564–1686, 2017.

[230] N. Weill, E. Therrien, V. Campagna-Slater, and N. Moitessier, "Methods for docking small molecules to macromolecules: A user's perspective. 1. the theory," *Current Pharmaceutical Design*, vol. 20, no. 20, pp. 3338–3359, 2014.

[231] V. Campagna-Slater, E. Therrien, N. Weill, and N. Moitessier, "Methods for docking small molecules to macromolecules: A user's perspective. 2. applications," *Current Pharmaceutical Design*, vol. 20, no. 20, pp. 3360–3372, 2014.

[232] X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, "Molecular docking: a powerful approach for structure-based drug discovery.," *Current Computer-Aided Drug Design*, vol. 7, no. 2, pp. 146–57, 2011.

[233] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson, "Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function," *Journal of Computational Chemistry*, vol. 19, no. 14, pp. 1639–1662, 1998.

[234] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and validation of a genetic algorithm for flexible docking," *Journal of Molecular Biology*, vol. 267, no. 3, pp. 727–748, 1997.

[235] C. M. Oshiro, I. D. Kuntz, and J. S. Dixon, "Flexible ligand docking using a genetic algorithm," *Journal of Computer-Aided Molecular Design*, vol. 9, no. 2, pp. 113–130, 1995.

[236] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor, "Improved protein-ligand docking using gold.," *Proteins*, vol. 52, no. 4, pp. 609–23, 2003.

[237] I. A. Guedes, C. S. de Magalhães, and L. E. Dardenne, "Receptor-ligand molecular docking," *Biophysical Reviews*, vol. 6, no. 1, pp. 75–87, 2014.

[238] F. Jiang and S. H. Kim, ""soft docking": Matching of molecular surface cubes," *Journal of Molecular Biology*, vol. 219, no. 1, pp. 79–102, 1991.

[239] D. A. Gschwend, C. Andrew, and I. D. Kuntz, "Molecular docking towards drug discovery," *Journal of Molecular Recognition*, vol. 9, no. October 1995, pp. 175–186, 1996.

[240] A. R. Leach, "Ligand docking to proteins with discrete side-chain flexibility," *Journal of Molecular Biology*, vol. 235, no. 1, pp. 345–356, 1994.

[241] J. Desmet, M. D. Maeyer, B. Hazes, and I. Lasters, "The dead-end elimination theorem and its use in protein side-chain positioning," *Nature*, vol. 356, no. 6369, pp. 539–542, 1992.

[242] D. S. Goodsell, H. Lauble, C. D. Stout, and A. J. Olson, "Automated docking in crystallography: Analysis of the substrates of aconitase," *Proteins: Structure, Function, and Bioinformatics*, vol. 17, no. 1, pp. 1–10, 1993.

[243] T. N. Hart and R. J. Read, "A multiple-start monte carlo docking method," *Proteins: Structure, Function, and Bioinformatics*, vol. 13, no. 3, pp. 206–222, 1992.

[244] A. Madadkar-Sobhani and V. Guallar, "Pele web server: atomistic study of biomolecular systems at your fingertips," *Nucleic Acids Research*, vol. 41, no. W1, pp. W322–W328, 2013.

[245] S.-Y. Huang and X. Zou, "Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking.," *Proteins*, vol. 66, no. 2, pp. 399–421, 2007.

[246] A. May and M. Zacharias, "Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: Evaluation on kinase inhibitor cross docking," *Journal of Medicinal Chemistry*, vol. 51, no. 12, pp. 3499–3506, 2008.

[247] M. Zacharias and H. Sklenar, "Harmonic modes as variables to approximately account for receptor flexibility in ligand-receptor docking simulations: Application to dna minor groove ligand complex," *Journal of Computational Chemistry*, vol. 20, no. 3, pp. 287–300, 1999.

[248] I. Kolossváry and W. C. Guida, "Low-mode conformational search elucidated: Application to c39h80 and flexible docking of 9-deazaguanine inhibitors into pnp," *Journal of Computational Chemistry*, vol. 20, no. 15, pp. 1671–1684, 1999.

[249] I. Kolossváry and G. M. Keserü, "Hessian-free low-mode conformational search for large-scale protein loop optimization: Application to c-jun n-terminal kinase jnk3," *Journal of Computational Chemistry*, vol. 22, no. 1, pp. 21–30, 2001.

[250] M. L. Teodoro, G. N. Phillips, and L. E. Kavraki, "Understanding protein flexibility through dimensionality reduction," *Journal of Computational Biology*, vol. 10, no. 3-4, pp. 617–634, 2003.

[251] M. L. Verdonk, J. C. Cole, and R. Taylor, "Superstar: A knowledge-based approach for identifying interaction sites in proteins," *Journal of Molecular Biology*, vol. 289, no. 4, pp. 1093–1108, 1999.

[252] R. Wang, Y. Lu, and S. Wang, "Comparative evaluation of 11 scoring functions for molecular docking," *Journal of Medicinal Chemistry*, vol. 46, no. 12, pp. 2287–2303, 2003.

[253] X. Huang, S. Y. and Zou, "Advances and challenges in protein-ligand docking," *International Journal of Molecular Sciences*, vol. 11, no. 8, pp. 3016–3034, 2010.

[254] P. Setny, "Water properties and potential of mean force for hydrophobic interactions of methane and nanoscopic pockets studied by computer simulations," *Journal of Chemical Physics*, vol. 127, no. 5, 2007.

[255] W. T. Mooij and M. L. Verdonk, "General and targeted statistical potentials for protein-ligand interactions," *Proteins: Structure, Function and Genetics*, vol. 61, no. 2, pp. 272–287, 2005.

[256] M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, and R. P. Mee, "Empirical scoring functions: I. the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes.," *Journal of Computer-aided Molecular Design*, vol. 11, no. 5, pp. 425–445, 1997.

[257] L. Verlet, "Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules," *Physical Review*, vol. 159, no. 1, pp. 98–103, 1967.

[258] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters," *Journal of Chemical Physics*, vol. 76, no. 1, pp. 637–649, 1982.

[259] S. Nosé, "A molecular dynamics method for simulations in the canonical ensemble," *Molecular Physics*, vol. 52, no. 2, pp. 255–268, 1984.

[260] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," *Physical Review A*, vol. 31, no. 3, pp. 1695–1697, 1985.

[261] B. Honig and M. Karplus, "Implications of torsional potential of retinal isomers for visual excitation," *Nature*, vol. 229, no. 5286, pp. 558–560, 1971.

[262] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," *Journal of Molecular Biology*, vol. 103, no. 2, pp. 227–249, 1976.

[263] S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, and M. J. Frisch, "A new oniom implementation in gaussian98. part i. the calculation of energies, gradients, vibrational frequencies and electric field derivatives," *Journal of Molecular Structure: THEOCHEM*, vol. 461-462, pp. 1–21, 1999.

[264] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, N. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. J. Throssell, A. Mont-

gomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, "Gaussian 09, revision a.02," 2016.

[265] L. W. Chung, W. M. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H. B. Li, L. Ding, and K. Morokuma, "The oniom method and its applications," 2015.

[266] V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret, and G. G. Ferenczy, "Quantum mechanical computations on very large molecular systems: The local self-consistent field method," *Journal of Computational Chemistry*, vol. 15, no. 3, pp. 269–282, 1994.

[267] N. Reuter, A. Dejaegere, B. Maigret, and M. Karplus, "Frontier bonds in qm/mm methods: A comparison of different approaches," *Journal of Physical Chemistry A*, vol. 104, no. 8, pp. 1720–1735, 2000.

[268] H. M. Senn and W. Thiel, "Qm/mm methods for biomolecular systems," *Angewandte Chemie - International Edition*, vol. 48, no. 7, pp. 1198–1229, 2009.

[269] C. V. Sumowski and C. Ochsenfeld, "A convergence study of qm/mm isomerization energies with the selected size of the qm region for peptidic systems," *Journal of Physical Chemistry A*, vol. 113, no. 43, pp. 11734–11741, 2009.

[270] P. E. M. Siegbahn and F. Himo, "The quantum chemical cluster approach for modeling enzyme reactions," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, no. 3, pp. 323–336, 2011.

[271] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. Vandevondele, "Cp2k: Atomistic simulations of condensed matter systems," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 4, no. 1, pp. 15–25, 2014.

[272] D. J. Price and C. L. Brooks III, "A modified tip3p water potential for simulation with ewald summation," *Journal of Chemical Physics*, vol. 121, no. 20, pp. 10096–10103, 2004.

[273] J. L. Abascal and C. Vega, "The water forcefield: Importance of dipolar and quadrupolar interactions," *Journal of Physical Chemistry C*, vol. 111, no. 43, pp. 15811–15822, 2007.

[274] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio Jr, *et al.*, "Current status of the amoeba polarizable force field," *Journal of Physical Chemistry B*, vol. 114, no. 8, pp. 2549–2564, 2010.

[275] E. Goldwaser, B. de Courcy, L. Demange, C. Garbay, F. Raynaud, R. Hadj-Slimane, J.-P. Piquemal, and N. Gresh, "Conformational analysis of a poly-conjugated protein-binding ligand by joint quantum chemistry and polarizable molecular mechanics. addressing the issues of anisotropy, conjugation, polarization, and multipole transferability," *Journal of Molecular Modeling*, vol. 20, no. 11, p. 2472, 2014.

[276] S. Y. Liem and P. L. Popelier, "The hydration of serine: multipole moments versus point charges," *Physical Chemistry Chemical Physics*, vol. 16, no. 9, pp. 4122–4134, 2014.

[277] H. Yu and W. F. van Gunsteren, "Charge-on-spring polarizable water models revisited: from water clusters to liquid water to ice," *Journal of Chemical Physics*, vol. 121, no. 19, pp. 9549–9564, 2004.

[278] M. Steele-MacInnis, J. Reimer, and S. Bachmann, "Hydrothermal properties of the cos/d2 water model: a polarizable charge-on-spring water model, at elevated temperatures and pressures," *RSC Advances*, vol. 5, no. 93, pp. 75846–75856, 2015.

[279] C. L. Brooks, B. M. Pettitt, and M. Karplus, "Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids," *Journal of Chemical Physics*, vol. 83, no. 11, pp. 5897–5908, 1985.

[280] M. Bergdorf, C. Peter, and P. H. Hünenberger, "Influence of cut-off truncation and artificial periodicity of electrostatic interactions in molecular simulations of solvated ions: A continuum electrostatics study," *Journal of Chemical Physics*, vol. 119, no. 17, pp. 9129–9144, 2003.

[281] P. J. Steinbach and B. R. Brooks, "New spherical???cutoff methods for long???range forces in macromolecular simulation," *Journal of Computational Chemistry*, vol. 15, no. 7, pp. 667–683, 1994.

[282] P. P. Ewald, "Die berechnung optischer und elektrostatischer gitterpotentiale," *Annalen der Physik*, vol. 369, no. 3, pp. 253–287, 1921.

[283] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh ewald method," *Journal of Chemical Physics*, vol. 103, no. 19, pp. 8577–8593, 1995.

[284] N. M. Fischer, P. J. Van Maaren, J. C. Ditz, A. Yildirim, and D. Van Der Spoel, "Properties of organic liquids when simulated with long-range lennard-jones interactions," *Journal of Chemical Theory and Computation*, vol. 11, no. 7, pp. 2938–2944, 2015.

[285] B. Mennucci and R. Cammi, *Continuum solvation models in chemical physics: from theory to applications*. John Wiley & Sons, 2008.

[286] A. V. Marenich, C. J. Cramer, and D. G. Truhlar, "Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions," *Journal of Physical Chemistry B*, vol. 113, no. 18, pp. 6378–6396, 2009.

[287] 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*, no. 5, pp. 799–805, 1993.

[288] K. A, "From quantum chemistry to fluid phase thermodynamics and drug design," *Journal of Chemical & Engineering Data*, vol. 51, no. 4, pp. 1480–1480, 2006.

[289] A. Klamt and V. Jonas, "Treatment of the outlying charge in continuum solvation models," *Journal of Chemical Physics*, vol. 105, no. 22, pp. 9972–9981, 1996.

[290] E. Michael, S. Polydorides, T. Simonson, and G. Archontis, "Simple models for nonpolar solvation: Parameterization and testing," *Journal of Computational Chemistry*, vol. 38, no. 29, pp. 2509–2519, 2017.

[291] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "Ucsf chimera - a visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.

[292] Y. Yang, M. N. Weaver, and K. M. Merz, "Assessment of the "6-31+gt; + lanl2dz" mixed basis set coupled with density functional theory methods and the effective core potential: Prediction of heats of formation and ionization potentials for first-row-transition-metal complexes," *Journal of Physical Chemistry A*, vol. 113, no. 36, pp. 9843–9851, 2009.

[293] N. Mardirossian and M. Head-Gordon, "Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals," *Molecular Physics*, vol. 115, no. 19, pp. 2315–2372, 2017.

[294] L. Yan, Y. Lu, and X. Li, "A density functional theory protocol for the calculation of redox potentials of copper complexes," *Physical Chemistry Chemical Physics*, vol. 18, no. 7, pp. 5529–5536, 2016.

[295] S. McAnanama-Brereton and M. P. Waller, "Rational density functional selection using game theory," *Journal of Chemical Information and Modeling*, p. acs.jcim.7b00542, 2017.

[296] K. E. Riley and K. M. Merz, "Assessment of density functional theory methods for the computation of heats of formation and ionization potentials of systems containing third row transition metals.," *Journal of Physical Chemistry A*, vol. 111, no. 27, pp. 6044–53, 2007.

[297] B. Seebeck, I. Reulecke, A. Kämper, and M. Rarey, "Modeling of metal interaction geometries for protein-ligand docking," *Proteins: Structure, Function and Genetics*, vol. 71, no. 3, pp. 1237–1254, 2008.

[298] G. Sciortino, J. Rodríguez-Guerra Pedregal, A. Lledós, E. Garribba, and J. D. Maréchal, "Prediction of the interaction of metallic moieties with proteins: An update for protein-ligand docking techniques," *Journal of Computational Chemistry*, vol. 39, no. 1, pp. 42–51, 2018.

[299] A. Sali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints.," *Journal of Molecular Biology*, vol. 234, no. 3, pp. 779–815, 1993.

[300] A. Šali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints," *Journal of Molecular Biology*, vol. 234, no. 3, pp. 779–815, 1993.

[301] J. Rodríguez-Guerra Pedregal, G. Sciortino, J. Guasp, M. Municoy, and J.-D. Maréchal, "Gaudimm: A modular multi-objective platform for molecular modeling," *Journal of Computational Chemistry*, vol. 38, no. 24, pp. 2118–2126, 2017.

[302] J. M. Seminario, "Calculation of intramolecular force fields from second-derivative tensors," *International Journal of Quantum Chemistry*, vol. 60, no. 7, pp. 1271–1277, 1996.

[303] P. Li and K. M. Merz, "Mcpb.py: A python based metal center parameter builder," *Journal of Chemical Information and Modeling*, vol. 56, no. 4, pp. 599–604, 2016.

[304] S. Zheng, Q. Tang, J. He, S. Du, S. Xu, C. Wang, Y. Xu, and F. Lin, "Vffdt: A new software for preparing amber force field parameters for metal-containing molecular systems," *Journal of Chemical Information and Modeling*, vol. 56, no. 4, pp. 811–818, 2016.

[305] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The resp model," *Journal of Physical Chemistry*, vol. 97, no. 40, pp. 10269–10280, 1993.

[306] A. Jakalian, D. B. Jack, and C. I. Bayly, "Fast, efficient generation of high-quality atomic charges. am1-bcc model: Ii. parameterization and validation," *Journal of Computational Chemistry*, vol. 23, no. 16, pp. 1623–1641, 2002.

[307] J. Li, T. Zhu, C. J. Cramer, and D. G. Truhlar, "New class iv charge model for extracting accurate partial charges from wave functions," *Journal of Physical Chemistry A*, vol. 102, no. 10, pp. 1820–1831, 1998.

[308] R. S. Mulliken, "Electronic population analysis on lcao-mo molecular wave functions. i," *Journal of Chemical Physics*, vol. 23, no. 10, pp. 1833–1840, 1955.

[309] J. Gasteiger and M. Marsili, "A new model for calculating atomic charges in molecules," *Tetrahedron Letters*, vol. 19, no. 34, pp. 3181–3184, 1978.

[310] J. Gasteiger and M. Marsili, "Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges," *Tetrahedron*, vol. 36, no. 22, pp. 3219–3228, 1980.

[311] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations," *Journal of Molecular Graphics and Modelling*, vol. 25, no. 2, pp. 247–260, 2006.

[312] J. Aaqvist and A. Warshel, "Calculations of free energy profiles for the staphylococcal nuclease catalyzed reaction," *Biochemistry*, vol. 28, no. 11, pp. 4680–4689, 1989.

[313] F. Duarte, P. Bauer, A. Barrozo, B. A. Amrein, M. Purg, J. Åqvist, and S. C. L. Kamerlin, "Force field independent metal parameters using a nonbonded dummy model," *Journal of Physical Chemistry B*, vol. 118, no. 16, pp. 4351–4362, 2014.

[314] A. Vedani and D. W. Huhta, "A new force field for modeling metalloproteins," *Journal of the American Chemical Society*, vol. 112, no. 12, pp. 4759–4767, 1990.

[315] N. Gresh, C. Policar, and C. Giessner-Prettre, "Modeling copper (i) complexes: Sibfa molecular mechanics versus ab initio energetics and geometrical arrangements," *Journal of Physical Chemistry A*, vol. 106, no. 23, pp. 5660–5670, 2002.

[316] V. J. Burton, R. J. Deeth, C. M. Kemp, and P. J. Gilbert, "Molecular mechanics for coordination complexes: the impact of adding d-electron stabilization energies," *Journal of the American Chemical Society*, vol. 117, no. 32, pp. 8407–8415, 1995.

[317] N. Gresh, J.-P. Piquemal, and M. Krauss, "Representation of zn (ii) complexes in polarizable molecular mechanics. further refinements of the electrostatic and short-range contributions. comparisons with parallel ab initio computations," *Journal of Computational Chemistry*, vol. 26, no. 11, pp. 1113–1130, 2005.

[318] B. De Courcy, J.-P. Piquemal, and N. Gresh, "Energy analysis of zn polycoordination in a metalloprotein environment and of the role of a neighboring aromatic residue. what is the impact of polarization?," *Journal of Chemical Theory and Computation*, vol. 4, no. 10, pp. 1659–1668, 2008.

[319] R. J. Deeth, "Comprehensive molecular mechanics model for oxidized type i copper proteins: active site structures, strain energies, and entatic bulging," *Inorganic Chemistry*, vol. 46, no. 11, pp. 4492–4503, 2007.

[320] A. Liese, K. Seelbach, and C. Wandrey, "Industrial biotransformations, 2nd completely revised and extended edition," 2006.

[321] J. Guo and P. Teo, "Anti-markovnikov oxidation and hydration of terminal olefins," *Dalton Transactions*, vol. 43, no. 19, pp. 6952–6964, 2014.

[322] T. Tokoroyama, "Discovery of the michael reaction," *European Journal of Organic Chemistry*, vol. 2010, no. 10, pp. 2009–2016, 2010.

[323] V. Resch and U. Hanefeld, "The selective addition of water," *Catalysis Science & Technology*, vol. 5, no. 3, pp. 1385–1399, 2015.

[324] O. Gawron and T. P. Fondy, "Stereochemistry of the fumarase and aspartase catalyzed reactions and of the krebs cycle from fumaric acid to d-isocitric acid1, 2," *Journal of the American Chemical Society*, vol. 81, no. 23, pp. 6333–6334, 1959.

[325] P. WILLADSEN and H. EGGERER, "Substrate stereochemistry of the enoyl-coa hydratase reaction," *The FEBS Journal*, vol. 54, no. 1, pp. 247–252, 1975.

[326] G. Agnihotri and H.-w. Liu, "Enoyl-coa hydratase: Reaction, mechanism, and inhibition," *Bioorganic & Medicinal Chemistry*, vol. 11, no. 1, pp. 9–20, 2003.

[327] K. Moerke, D. Cloutier, B. Lane, E. Person, T. Onasch, *et al.*, "Importance of historical contingency in the stereochemistry of hydratase-dehydratase enzymes," *Science*, vol. 269, no. 5223, pp. 527–529, 1995.

[328] J. Jin and U. Hanefeld, "The selective addition of water to c [double bond, length as m-dash] c bonds; enzymes are the best chemists," *Chemical Communications*, vol. 47, no. 9, pp. 2502–2510, 2011.

[329] X. Sheng and F. Himo, "Theoretical study of enzyme promiscuity: Mechanisms of hydration and carboxylation activities of phenolic acid decarboxylase," *ACS Catalysis*, vol. 7, no. 3, pp. 1733–1741, 2017.

[330] C. Wuensch, J. Gross, G. Steinkellner, K. Gruber, S. M. Glueck, and K. Faber, "Asymmetric enzymatic hydration of hydroxystyrene derivatives," *Angewandte Chemie - International Edition*, vol. 52, no. 8, pp. 2293–2297, 2013.

[331] J. Jin, P. C. Oskam, S. K. Karmee, A. J. Straathof, and U. Hanefeld, "Mhyadh catalysed michael addition of water and in situ oxidation," *Chemical Communications*, vol. 46, no. 45, pp. 8588–8590, 2010.

[332] R. P. Megens and G. Roelfes, "Dna-based catalytic enantioselective intermolecular oxa-michael addition reactions," *Chemical Communications*, vol. 48, no. 51, pp. 6366–6368, 2012.

[333] B.-S. Chen, L. G. Otten, and U. Hanefeld, "Stereochemistry of enzymatic water addition to c= c bonds," *Biotechnology Advances*, vol. 33, no. 5, pp. 526–546, 2015.

[334] R.-Z. Liao, J.-G. Yu, and F. Himo, "Mechanism of tungsten-dependent acetylene hydratase from quantum chemical calculations," *Proceedings of the National Academy of Sciences*, vol. 107, no. 52, pp. 22523–22527, 2010.

[335] G. Petersson, A. Bennett, T. G. Tensfeldt, M. A. Al-Laham, W. A. Shirley, and J. Mantzaris, "A complete basis set model chemistry. i. the total energies of closed-shell atoms and hydrides of the first-row elements," *Journal of Chemical Physics*, vol. 89, no. 4, pp. 2193–2218, 1988.

[336] G. Petersson and M. A. Al-Laham, "A complete basis set model chemistry. ii. open-shell systems and the total energies of the first-row atoms," *Journal of Chemical Physics*, vol. 94, no. 9, pp. 6081–6090, 1991.

[337] V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern, and L. A. Curtiss, "6-31g* basis set for third-row atoms," *Journal of Computational Chemistry*, vol. 22, no. 9, pp. 976–984, 2001.

[338] M. Dolg, U. Wedig, H. Stoll, and H. Preuss, "Energy-adjusted abinitio pseudopotentials for the first row transition elements," *The Journal of Chemical Physics*, vol. 86, no. 2, pp. 866–872, 1987.

[339] S. Genheden and U. Ryde, "Comparison of end-point continuum-solvation methods for the calculation of protein–ligand binding free energies," *Proteins: Structure, Function, and Bioinformatics*, vol. 80, no. 5, pp. 1326–1342, 2012.

[340] A. Stirling and I. Pápai, "H2co3 forms via hco3- in water," *Journal of Physical Chemistry B*, vol. 114, no. 50, pp. 16854–16859, 2010.

[341] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, "Amber, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules," *Computer Physics Communications*, vol. 91, no. 1-3, pp. 1–41, 1995.

[342] M. Crowley, T. Darden, T. Cheatham, and D. Deerfield, "Adventures in improving the scaling and accuracy of a parallel molecular dynamics program," *Journal of Supercomputing*, vol. 11, no. 3, pp. 255–278, 1997.

[343] "The cp2k developers group." `http://www.cp2k.org/`. Accessed: 2013-05-31.

[344] J. VandeVondele and J. Hutter, "Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases," *Journal of Chemical Physics*, vol. 127, no. 11, p. 114105, 2007.

[345] H. Stoll, P. Fuentealba, P. Schwerdtfeger, J. v. Flad, L. v. Szentpaly, and H. Preuss, "Cu and ag as one-valence-electron atoms: Ci results and quadrupole corrections for cu2, ag2, cuh, and agh," *Journal of Chemical Physics*, vol. 81, no. 6, pp. 2732–2736, 1984.

[346] R. L. Dunbrack Jr, "Rotamer libraries in the 21st century," *Current Opinion in Structural Biology*, vol. 12, no. 4, pp. 431–440, 2002.

[347] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *Journal of Chemical Physics*, vol. 79, no. 2, pp. 926–935, 1983.

[348] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *Journal of Computational Physics*, vol. 23, no. 3, pp. 327–341, 1977.

[349] T. Schneider and E. Stoll, "Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions," *Physical Review B*, vol. 17, no. 3, p. 1302, 1978.

[350] A. Brünger, C. L. Brooks III, and M. Karplus, "Stochastic boundary conditions for molecular dynamics simulations of st2 water," *Chemical Physics Letters*, vol. 105, no. 5, pp. 495–500, 1984.

[351] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid monte carlo," *Physics Letters B*, vol. 195, no. 2, pp. 216–222, 1987.

[352] P. Eastman and V. Pande, "Openmm: a hardware-independent framework for molecular simulations," *Computing in Science & Engineering*, vol. 12, no. 4, pp. 34–39, 2010.

[353] B. J. Bahnson, V. E. Anderson, and G. A. Petsko, "Structural mechanism of enoyl-coa hydratase: three atoms from a single water are added in either an e1cb stepwise or concerted fashion," *Biochemistry*, vol. 41, no. 8, pp. 2621–2629, 2002.

[354] J. P. Bennett, L. Bertin, B. Moulton, I. J. Fairlamb, A. M. Brzozowski, N. J. Walton, and G. Grogan, "A ternary complex of hydroxycinnamoyl-coa

hydratase–lyase (hchl) with acetyl-coa and vanillin gives insights into substrate specificity and mechanism," *Biochemical Journal*, vol. 414, no. 2, pp. 281–289, 2008.

[355] P. Kasaragod, W. Schmitz, J. K. Hiltunen, and R. K. Wierenga, "The isomerase and hydratase reaction mechanism of the crotonase active site of the multifunctional enzyme (type-1), as deduced from structures of complexes with 3s-hydroxy-acyl-coa," *The FEBS Journal*, vol. 280, no. 13, pp. 3160–3175, 2013.

[356] C. Letondor, N. Humbert, and T. R. Ward, "Artificial metalloenzymes based on biotin-avidin technology for the enantioselective reduction of ketones by transfer hydrogenation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 13, pp. 4683–4687, 2005.

[357] H. Mallin, M. Hestericová, R. Reuter, and T. R. Ward, "Library design and screening protocol for artificial metalloenzymes based on the biotin-streptavidin technology," *Nature Protocols*, vol. 11, no. 5, p. 835, 2016.

[358] R. L. Dunbrack Jr and M. Karplus, "Backbone-dependent rotamer library for proteins application to side-chain prediction," *Journal of Molecular Biology*, vol. 230, no. 2, pp. 543–574, 1993.

[359] F. Neese, "The orca program system," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, no. 1, pp. 73–78, 2012.

[360] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, "Scalable molecular dynamics with namd," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1781–1802, 2005.

[361] A. Onufriev, D. Bashford, and D. A. Case, "Exploring protein native states and large-scale conformational changes with a modified generalized born model,"

Proteins: Structure, Function, and Bioinformatics, vol. 55, no. 2, pp. 383–394, 2004.

[362] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, *et al.*, "Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models," *Accounts of Chemical Research*, vol. 33, no. 12, pp. 889–897, 2000.

[363] K. Liu and H. Kokubo, "Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations: A cross-docking study," *Journal of Chemical Information and Modeling*, vol. 57, no. 10, pp. 2514–2522, 2017.

[364] J. Shen, W. Zhang, H. Fang, R. Perkins, W. Tong, and H. Hong, "Homology modeling, molecular docking, and molecular dynamics simulations elucidated α-fetoprotein binding modes," *BMC Bioinformatics*, vol. 14, no. 14, p. S6, 2013.

[365] C. Severin Lupala, B. Rasaeifar, P. Gomez Gutierrez, and J. J. Pérez González, "Using molecular dynamics for the refinement of atomistic models of gpcrs by homology modeling," *Journal of Biomolecular Structure and Dynamics*, pp. 1–13, 2017.

[366] D. Gioia, M. Bertazzo, M. Recanatini, M. Masetti, and A. Cavalli, "Dynamic docking: A paradigm shift in computational drug discovery," *Molecules*, vol. 22, no. 11, p. 2029, 2017.

[367] M. A. Isa, R. S. Majumdhar, and S. Haider, "In silico docking and molecular dynamics simulation of 3-dehydroquinate synthase (dhqs) from mycobacterium tuberculosis," *Journal of Molecular Modeling*, vol. 24, no. 6, p. 132, 2018.

[368] W. J. Song and F. A. Tezcan, "A designed supramolecular protein assembly with in vivo enzymatic activity," *Science*, vol. 346, no. 6216, pp. 1525–1528, 2014.

[369] H.-S. Park, S.-H. Nam, J. K. Lee, C. N. Yoon, B. Mannervik, S. J. Benkovic, and H.-S. Kim, "Design and evolution of new catalytic activity with an existing protein scaffold," *Science*, vol. 311, no. 5760, pp. 535–538, 2006.

[370] P. K. Madoori, H. Agustiandari, A. J. Driessen, and A.-M. W. Thunnissen, "Structure of the transcriptional regulator lmrr and its mechanism of multidrug recognition," *The EMBO Journal*, vol. 28, no. 2, pp. 156–166, 2009.

[371] J. Xie, W. Liu, and P. G. Schultz, "A genetically encoded bidentate, metal-binding amino acid," *Angewandte Chemie - International Edition*, vol. 119, no. 48, pp. 9399–9402, 2007.

[372] C. F. Nising and S. Braese, "Recent developments in the field of oxa-michael reactions," *Chemical Society Reviews*, vol. 41, no. 3, pp. 988–999, 2012.

[373] M. Gongora-Benitez, J. Tulla-Puche, and F. Albericio, "Multifaceted roles of disulfide bonds. peptides as therapeutics," *Chemical Reviews*, vol. 114, no. 2, pp. 901–926, 2013.

[374] B. Gomes, M. T. Augusto, M. R. Felício, A. Hollmann, O. L. Franco, S. Gonçalves, and N. C. Santos, "Designing improved active peptides for therapeutic approaches against infectious diseases," *Biotechnology Advances*, 2018.

[375] E. M. Driggers, S. P. Hale, J. Lee, and N. K. Terrett, "The exploration of macrocycles for drug discovery—an underexploited structural class," *Nature Reviews Drug Discovery*, vol. 7, no. 7, p. 608, 2008.

[376] P. Thapa, M. J. Espiritu, C. Cabalteja, and J.-P. Bingham, "The emergence of cyclic peptides: The potential of bioengineered peptide drugs," *International Journal of Peptide Research and Therapeutics*, vol. 20, no. 4, pp. 545–551, 2014.

[377] K. Fosgerau and T. Hoffmann, "Peptide therapeutics: current status and future directions," *Drug Discovery Today*, vol. 20, no. 1, pp. 122–128, 2015.

[378] D. J. Craik, D. P. Fairlie, S. Liras, and D. Price, "The future of peptide-based drugs," *Chemical Biology & Drug Design*, vol. 81, no. 1, pp. 136–147, 2013.

[379] A. Aguzzi and T. O'connor, "Protein aggregation diseases: pathogenicity and therapeutic perspectives," *Nature Reviews Drug Discovery*, vol. 9, no. 3, p. 237, 2010.

[380] S. Learte-Aymamí, N. Curado, J. Rodríguez, M. E. Vázquez, and J. L. Mascareñas, "Metal-dependent dna recognition and cell internalization of designed, basic peptides," *Journal of the American Chemical Society*, vol. 139, no. 45, pp. 16188–16193, 2017.

[381] C. Exley, "Aluminium and iron, but neither copper nor zinc, are key to the precipitation of β-sheets of aβ 42 in senile plaque cores in alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 10, no. 2-3, pp. 173–177, 2006.

[382] M. Mold, L. Ouro-Gnao, B. M. Wieckowski, and C. Exley, "Copper prevents amyloid-β 1–42 from forming amyloid fibrils under near-physiological conditions in vitro," *Scientific Reports*, vol. 3, p. 1256, 2013.

[383] C. Exley, M. Mold, E. Shardlow, B. Shuker, B. Ikpe, L. Wu, and P. E. Fraser, "Copper is a potent inhibitor of the propensity for human proiapp1-48 to form amyloid fibrils in vitro," *Journal of Diabetes Research and Clinical Metabolism*, vol. 1, no. 1, p. 3, 2012.

[384] L. Pickart, J. M. Vasquez-Soltero, and A. Margolina, "The effect of the human peptide ghk on gene expression relevant to nervous system function and cognitive decline," *Brain Sciences*, vol. 7, no. 2, p. 20, 2017.

[385] Z. T. Ball, "Designing enzyme-like catalysts: a rhodium (ii) metallopeptide case study," *Accounts of Chemical Research*, vol. 46, no. 2, pp. 560–570, 2012.

[386] H. Wennemers, "Asymmetric catalysis with peptides," *Chemical Communications*, vol. 47, no. 44, pp. 12036–12041, 2011.

[387] F. Kolundzic, M. N. Noshi, M. Tjandra, M. Movassaghi, and S. J. Miller, "Chemoselective and enantioselective oxidation of indoles employing aspartyl peptide catalysts," *Journal of the American Chemical Society*, vol. 133, no. 23, pp. 9104–9111, 2011.

[388] K. W. Fiori, A. L. Puchlopek, and S. J. Miller, "Enantioselective sulfonylation reactions mediated by a tetrapeptide catalyst," *Nature Chemistry*, vol. 1, no. 8, p. 630, 2009.

[389] R. H. Kimura, A. M. Levin, F. V. Cochran, and J. R. Cochran, "Engineered cystine knot peptides that bind $\alpha v\beta 3$, $\alpha v\beta 5$, and $\alpha 5\beta 1$ integrins with low-nanomolar affinity," *Proteins: Structure, Function, and Bioinformatics*, vol. 77, no. 2, pp. 359–369, 2009.

[390] P. D. Renfrew, E. J. Choi, R. Bonneau, and B. Kuhlman, "Incorporation of noncanonical amino acids into rosetta and use in computational protein-peptide interface design," *PLoS One*, vol. 7, no. 3, p. e32637, 2012.

[391] S. Chakraborty, J. Yudenfreund Kravitz, P. W. Thulstrup, L. Hemmingsen, W. F. DeGrado, and V. L. Pecoraro, "Design of a three-helix bundle capable of binding heavy metals in a triscysteine environment," *Angewandte Chemie - International Edition*, vol. 50, no. 9, pp. 2049–2053, 2011.

[392] I. V. Korendovych, A. Senes, Y. H. Kim, J. D. Lear, H. C. Fry, M. J. Therien, J. K. Blasie, F. A. Walker, and W. F. DeGrado, "De novo design and molecular

assembly of a transmembrane diporphyrin-binding protein complex," *Journal of the American Chemical Society*, vol. 132, no. 44, pp. 15516–15518, 2010.

[393] O. Maglio, F. Nastri, J. R. Calhoun, S. Lahr, H. Wade, V. Pavone, W. F. DeGrado, and A. Lombardi, "Artificial di-iron proteins: solution characterization of four helix bundles containing two distinct types of inter-helical loops," *JBIC Journal of Biological Inorganic Chemistry*, vol. 10, no. 5, pp. 539–549, 2005.

[394] R. E. Sharp, J. R. Diers, D. F. Bocian, and P. L. Dutton, "Differential binding of iron (iii) and zinc (ii) protoporphyrin ix to synthetic four-helix bundles," *Journal of the American Chemical Society*, vol. 120, no. 28, pp. 7103–7104, 1998.

[395] E. N. Salgado, R. J. Radford, and F. A. Tezcan, "Metal-directed protein self-assembly," *Accounts of Chemical Research*, vol. 43, no. 5, pp. 661–672, 2010.

[396] B. R. Lichtenstein, T. A. Farid, G. Kodali, L. A. Solomon, J. R. Anderson, M. M. Sheehan, N. M. Ennist, B. A. Fry, S. E. Chobot, C. Bialas, *et al.*, "Engineering oxidoreductases: maquette proteins designed from scratch," *Biochemical Society Transactions*, vol. 40, no. 3, pp. 561–566, 2012.

[397] A. K. Jones, B. R. Lichtenstein, A. Dutta, G. Gordon, and P. L. Dutton, "Synthetic hydrogenases: incorporation of an iron carbonyl thiolate into a designed peptide," *Journal of the American Chemical Society*, vol. 129, no. 48, pp. 14844–14845, 2007.

[398] A. M. Fernández-Escamilla, S. Ventura, L. Serrano, and M. Jiménez, "Design and nmr conformational study of a β-sheet peptide based on betanova and ww domains," *Protein Science*, vol. 15, no. 10, pp. 2278–2289, 2006.

[399] M. J. Macias, V. Gervais, C. Civera, and H. Oschkinat, "Structural analysis of ww domains and design of a ww prototype," *Nature Structural and Molecular Biology*, vol. 7, no. 5, p. 375, 2000.

[400] A. L. Stewart, J. H. Park, and M. L. Waters, "Redesign of a ww domain peptide for selective recognition of single-stranded dna," *Biochemistry*, vol. 50, no. 13, pp. 2575–2584, 2011.

[401] Y. Kato, Y. Hino, K. Nagata, and M. Tanokura, "Solution structure and binding specificity of fbp11/hypa ww domain as group-ii/iii," *Proteins: Structure, Function, and Bioinformatics*, vol. 63, no. 1, pp. 227–234, 2006.

[402] J. Habazettl, A. Reiner, and T. Kiefhaber, "Nmr structure of a monomeric intermediate on the evolutionarily optimized assembly pathway of a small trimerization domain," *Journal of Molecular Biology*, vol. 389, no. 1, pp. 103–114, 2009.

[403] W. Chen, S. Enck, J. L. Price, D. L. Powers, E. T. Powers, C.-H. Wong, H. J. Dyson, and J. W. Kelly, "Structural and energetic basis of carbohydrate–aromatic packing interactions in proteins," *Journal of the American Chemical Society*, vol. 135, no. 26, pp. 9877–9884, 2013.

[404] R. Zhou, G. G. Maisuradze, D. Suñol, T. Todorovski, M. J. Macias, Y. Xiao, H. A. Scheraga, C. Czaplewski, and A. Liwo, "Folding kinetics of ww domains with the united residue force field for bridging microscopic motions and experimental measurements," *Proceedings of the National Academy of Sciences*, vol. 111, no. 51, pp. 18243–18248, 2014.

[405] C. Appelt, A. Wessolowski, M. Dathe, and P. Schmieder, "Structures of cyclic, antimicrobial peptides in a membrane-mimicking environment define require-

ments for activity," *Journal of Peptide Science*, vol. 14, no. 4, pp. 524–527, 2008.

[406] S. Schlee, T. Klein, M. Schumacher, J. Nazet, R. Merkl, H.-J. Steinhoff, and R. Sterner, "Relationship of catalysis and active site loop dynamics in the ($\beta\alpha$) 8-barrel enzyme indole-3-glycerol phosphate synthase," *Biochemistry*, 2018.

[407] N. Kamariah, B. Eisenhaber, F. Eisenhaber, and G. Grüber, "Active site cp-loop dynamics modulate substrate binding, catalysis, oligomerization, stability, over-oxidation and recycling of 2-cys peroxiredoxins," *Free Radical Biology and Medicine*, vol. 118, pp. 59–70, 2018.

[408] B. Sharma, S. N. Jamdar, B. Ghosh, P. Yadav, A. Kumar, S. Kundu, V. D. Goyal, and R. D. Makde, "Active site gate of m32 carboxypeptidases illuminated by crystal structure and molecular dynamics simulations," *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, vol. 1865, no. 11, pp. 1406–1415, 2017.

[409] K. A. Wilson and S. D. Wetmore, "Conformational flexibility of the benzylguanine adduct in a bypass polymerase active site permits replication: Insights from molecular dynamics simulations," *Chemical Research in Toxicology*, vol. 30, no. 11, pp. 2013–2022, 2017.

[410] M. K. Rout, B. L. Lee, A. Lin, W. Xiao, and L. Spyracopoulos, "Active site gate dynamics modulate the catalytic activity of the ubiquitination enzyme e2-25k," *Scientific Reports*, vol. 8, no. 1, p. 7002, 2018.

[411] K. Sen, S. Horrell, D. Kekilli, C. W. Yong, T. W. Keal, H. Atakisi, D. W. Moreau, R. E. Thorne, M. A. Hough, and R. W. Strange, "Active-site protein dynamics and solvent accessibility in native achromobacter cycloclastes copper nitrite reductase," *IUCrJ*, vol. 4, no. 4, 2017.

[412] D. Dutta and S. Mishra, "Active site dynamics in substrate hydrolysis catalyzed by dape enzyme and its mutants from hybrid qm/mm-molecular dynamics simulation," *Journal of Physical Chemistry B*, vol. 121, no. 29, pp. 7075–7085, 2017.

[413] R. A. Luirink, S. J. Dekker, L. Capoferri, L. F. Janssen, C. L. Kuiper, M. E. Ari, N. P. Vermeulen, J. C. Vos, J. N. Commandeur, and D. P. Geerke, "A combined computational and experimental study on selective flucloxacillin hydroxylation by cytochrome p450 bm3 variants," *Journal of Inorganic Biochemistry*, 2018.

[414] R. H. Wilson, S. Zamfir, and I. Sumner, "Molecular dynamics simulations reveal a new role for a conserved active site asparagine in a ubiquitin-conjugating enzyme," *Journal of Molecular Graphics and Modelling*, vol. 76, pp. 403–411, 2017.

[415] X. Jiang, Y. Wang, L. Xu, G. Chen, and L. Wang, "Substrate binding interferes with active site conformational dynamics in endoglucanase cel5a from thermobifida fusca," *Biochemical and Biophysical Research Communications*, vol. 491, no. 1, pp. 236–240, 2017.

[416] R. Callender and R. B. Dyer, "The dynamical nature of enzymatic catalysis," *Accounts of Chemical Research*, vol. 48, no. 2, pp. 407–413, 2014.

[417] G. G. Hammes, S. J. Benkovic, and S. Hammes-Schiffer, "Flexibility, diversity, and cooperativity: pillars of enzyme catalysis," *Biochemistry*, vol. 50, no. 48, pp. 10422–10430, 2011.

[418] S. Hammes-Schiffer and S. J. Benkovic, "Relating protein motion to catalysis," *Annual Review of Biochemistry*, vol. 75, pp. 519–541, 2006.

[419] K. Henzler-Wildman and D. Kern, "Dynamic personalities of proteins," *Nature*, vol. 450, no. 7172, p. 964, 2007.

[420] Z. D. Nagel and J. P. Klinman, "A 21 st century revisionist's view at a turning point in enzymology," *Nature Chemical Biology*, vol. 5, no. 8, p. 543, 2009.

[421] C. K. Prier and F. H. Arnold, "Chemomimetic biocatalysis: exploiting the synthetic potential of cofactor-dependent enzymes to create new catalysts," *Journal of the American Chemical Society*, vol. 137, no. 44, pp. 13992–14006, 2015.

[422] O. F. Brandenberg, R. Fasan, and F. H. Arnold, "Exploiting and engineering hemoproteins for abiological carbene and nitrene transfer reactions," *Current Opinion in Biotechnology*, vol. 47, pp. 102–111, 2017.

[423] P. S. Coelho, E. M. Brustad, A. Kannan, and F. H. Arnold, "Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome p450 enzymes," *Science*, p. 1231434, 2012.

[424] P. S. Coelho, Z. J. Wang, M. E. Ener, S. A. Baril, A. Kannan, F. H. Arnold, and E. M. Brustad, "A serine-substituted p450 catalyzes highly efficient carbene transfer to olefins in vivo," *Nature Chemical Biology*, vol. 9, no. 8, p. 485, 2013.

[425] M. Bordeaux, V. Tyagi, and R. Fasan, "Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts," *Angewandte Chemie - International Edition*, vol. 127, no. 6, pp. 1764–1768, 2015.

[426] J. G. Gober, A. E. Rydeen, E. J. Gibson-O'Grady, J. B. Leuthaeuser, J. S. Fetrow, and E. M. Brustad, "Mutating a highly conserved residue in diverse cytochrome p450s facilitates diastereoselective olefin cyclopropanation," *ChemBioChem*, vol. 17, no. 5, pp. 394–397, 2016.

[427] A. Tinoco, V. Steck, V. Tyagi, and R. Fasan, "Highly diastereo-and enantioselective synthesis of trifluoromethyl-substituted cyclopropanes via myoglobin-

catalyzed transfer of trifluoromethylcarbene," *Journal of the American Chemical Society*, vol. 139, no. 15, pp. 5293–5296, 2017.

[428] V. Tyagi and R. Fasan, "Myoglobin-catalyzed olefination of aldehydes," *Angewandte Chemie - International Edition*, vol. 128, no. 7, pp. 2558–2562, 2016.

[429] M. J. Weissenborn, S. A. Löw, N. Borlinghaus, M. Kuhn, S. Kummer, F. Rami, B. Plietker, and B. Hauer, "Enzyme-catalyzed carbonyl olefination by the e. coli protein yfex in the absence of phosphines," *ChemCatChem*, vol. 8, no. 9, pp. 1636–1640, 2016.

[430] Z. J. Wang, N. E. Peck, H. Renata, and F. H. Arnold, "Cytochrome p450-catalyzed insertion of carbenoids into n–h bonds," *Chemical Science*, vol. 5, no. 2, pp. 598–601, 2014.

[431] G. Sreenilayam and R. Fasan, "Myoglobin-catalyzed intermolecular carbene n–h insertion with arylamine substrates," *Chemical Communications*, vol. 51, no. 8, pp. 1532–1534, 2015.

[432] S. J. Kan, R. D. Lewis, K. Chen, and F. H. Arnold, "Directed evolution of cytochrome c for carbon–silicon bond formation: Bringing silicon to life," *Science*, vol. 354, no. 6315, pp. 1048–1051, 2016.

[433] S. J. Kan, X. Huang, Y. Gumulya, K. Chen, and F. H. Arnold, "Genetically programmed chiral organoborane synthesis," *Nature*, vol. 552, no. 7683, p. 132, 2017.

[434] H. M. Key, P. Dydio, D. S. Clark, and J. F. Hartwig, "Abiological catalysis by artificial haem proteins containing noble metals in place of iron," *Nature*, vol. 534, no. 7608, p. 534, 2016.

[435] H. M. Key, P. Dydio, Z. Liu, J. Y.-E. Rha, A. Nazarenko, V. Seyedkazemi, D. S. Clark, and J. F. Hartwig, "Beyond iron: iridium-containing p450 enzymes for selective cyclopropanations of structurally diverse alkenes," *ACS Central Science*, vol. 3, no. 4, pp. 302–308, 2017.

[436] G. Sreenilayam, E. J. Moore, V. Steck, and R. Fasan, "Metal substitution modulates the reactivity and extends the reaction scope of myoglobin carbene transfer catalysts," *Advanced Synthesis & Catalysis*, vol. 359, no. 12, pp. 2076–2089, 2017.

[437] K. Oohora, H. Meichin, L. Zhao, M. W. Wolf, A. Nakayama, J.-y. Hasegawa, N. Lehnert, and T. Hayashi, "Catalytic cyclopropanation by myoglobin reconstituted with iron porphycene: Acceleration of catalysis due to rapid formation of the carbene species," *Journal of the American Chemical Society*, vol. 139, no. 48, pp. 17265–17268, 2017.

[438] G. Sreenilayam, E. J. Moore, V. Steck, and R. Fasan, "Stereoselective olefin cyclopropanation under aerobic conditions with an artificial enzyme incorporating an iron-chlorin e6 cofactor," *ACS Catalysis*, vol. 7, no. 11, pp. 7629–7633, 2017.

[439] T. Yonetani and T. Asakura, "Studies on cytochrome c peroxidase xv. comparison of manganese porphyrin-containing cytochrome c peroxidase, horseradish peroxidase, and myoglobin," *Journal of Biological Chemistry*, vol. 244, no. 17, pp. 4580–4588, 1969.

[440] T. Hayashi, H. Dejima, T. Matsuo, H. Sato, D. Murata, and Y. Hisaeda, "Blue myoglobin reconstituted with an iron porphycene shows extremely high oxygen affinity," *Journal of the American Chemical Society*, vol. 124, no. 38, pp. 11226–11227, 2002.

[441] C. J. Reedy and B. R. Gibney, "Heme protein assemblies," *Chemical Reviews*, vol. 104, no. 2, pp. 617–650, 2004.

[442] Y.-W. Lin, E. B. Sawyer, and J. Wang, "Rational heme protein design: All roads lead to rome," *Chemistry–An Asian Journal*, vol. 8, no. 11, pp. 2534–2544, 2013.

[443] C. Moser, M. Sheehan, N. Ennist, G. Kodali, C. Bialas, M. Englander, B. Discher, and P. Dutton, "De novo construction of redox active proteins," in *Methods in Enzymology*, vol. 580, pp. 365–388, Elsevier, 2016.

[444] K. Takeuchi, Y. Tokunaga, M. Imai, H. Takahashi, and I. Shimada, "Dynamic multidrug recognition by multidrug transcriptional repressor lmrr," *Scientific Reports*, vol. 4, p. 6922, 2014.

[445] D. Case, J. Berryman, R. Betz, D. Cerutti, T. Cheatham Iii, T. Darden, R. Duke, T. Giese, H. Gohlke, A. Goetz, *et al.*, "Amber 2015," *University of California, San Francisco*, 2015.

[446] C. J. Wilson, D. Apiyo, and P. Wittung-Stafshede, "Role of cofactors in metalloprotein folding," *Quarterly Reviews of Biophysics*, vol. 37, no. 3-4, pp. 285–314, 2004.

[447] V. Muñoz Robles, E. Ortega-Carrasco, E. González Fuentes, A. Lledós, and J.-D. Maréchal, "What can molecular modelling bring to the design of artificial inorganic cofactors?," *Faraday Discussions*, vol. 148, no. 0, pp. 137–159; discussion 207–228, 2011.

[448] Y. Watanabe, S. Fukuyoshi, M. Hiratsuka, N. Yamaotsu, S. Hirono, O. Takahashi, and A. Oda, "Prediction of three-dimensional structures and structural flexibilities of wild-type and mutant cytochrome p450 1a2 using molecular dy-

namics simulations," *Journal of Molecular Graphics and Modelling*, vol. 68, pp. 48–56, 2016.

[449] R. Hussain, I. Kumari, S. Sharma, M. Ahmed, T. A. Khan, and Y. Akhter, "Catalytic diversity and homotropic allostery of two cytochrome p450 monooxygenase like proteins from trichoderma brevicompactum," *JBIC Journal of Biological Inorganic Chemistry*, vol. 22, no. 8, pp. 1197–1209, 2017.

[450] L. Yang, Å. A. Skjevik, W.-G. H. Du, L. Noodleman, R. C. Walker, and A. W. Götz, "Water exit pathways and proton pumping mechanism in b-type cytochrome c oxidase from molecular dynamics simulations," *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1857, no. 9, pp. 1594–1606, 2016.

[451] L. Agulló, I. Buch, H. Gutiérrez-de Terán, D. Garcia-Dorado, and J. Villà-Freixa, "Computational exploration of the binding mode of heme-dependent stimulators into the active catalytic domain of soluble guanylate cyclase," *Proteins: Structure, Function, and Bioinformatics*, vol. 84, no. 10, pp. 1534–1548, 2016.

[452] M. Soloviov and M. Meuwly, "Co-dynamics in the active site of cytochrome c oxidase," *Journal of Chemical Physics*, vol. 140, no. 14, p. 04B605_1, 2014.

[453] Y.-W. Lin, Y.-M. Wu, and L.-F. Liao, "Dynamics comparison of two myoglobins with a distinct heme active site," *Journal of Molecular Modeling*, vol. 18, no. 4, pp. 1591–1596, 2012.

[454] G. Sciortino, J. Rodríguez-Guerra Pedregal, A. Lledós, E. Garribba, and J.-D. Maréchal, "Prediction of the interaction of metallic moieties with proteins: An update for protein-ligand docking techniques," *Journal of Computational Chemistry*, vol. 39, no. 1, pp. 42–51, 2018.

[455] J. Rodríguez-Guerra Pedregal, G. Sciortino, J. Guasp, M. Municoy, and J.-D. Maréchal, "Gaudimm: A modular multi-objective platform for molecular modeling," *Journal of Computational Chemistry*, vol. 38, no. 24, pp. 2118–2126, 2017.

[456] P. Arnoux, R. Haser, N. Izadi-Pruneyre, A. Lecroisey, and M. Czjzek, "Functional aspects of the heme bound hemophore hasa by structural analysis of various crystal forms," *Proteins: Structure, Function, and Bioinformatics*, vol. 41, no. 2, pp. 202–210, 2000.

[457] N. Izadi, Y. Henry, J. Haladjian, M. E. Goldberg, C. Wandersman, M. Delepierre, and A. Lecroisey, "Purification and characterization of an extracellular heme-binding protein, hasa, involved in heme iron acquisition," *Biochemistry*, vol. 36, no. 23, pp. 7050–7057, 1997.