SAPIENZA
UNIVERSITÀ DI ROMA

Hybrid Computational Approaches for the Elucidation of Complex Biocatalytic Mechanisms

Candidato:

Matteo Farina

Relatori:

Prof. Enrico Bodo (Tutor)
Prof.ssa Isabella Daidone (Co-tutor)

Corso:

Dottorato di ricerca in Scienze Chimiche

Scuola di Dottorato Vito Volterra
XXXVIII Ciclo

Abstract

We report two mechanistic investigations on recently discovered, industrially relevant biocatalytic systems conducted using a hybrid computational approach integrating Quantum Mechanics/Molecular Mechanics (QM/MM) with the Molecular Dynamics/Perturbed Matrix Method (MD-PMM). In the first study, concerning the reduction of 2-pentanone catalyzed by the Horse Liver Alcohol Dehydrogenase (HLADH) in the presence of the biomimetic cofactor, benzyl-1,4-dihydronicotinamide (BNAH), we propose a revised mechanism that follows a sequential pathway in which an endothermic hydride transfer (HT) precedes a strongly exothermic proton transfer (PT).

The Gibbs free-energy barrier for HT, calculated using MD-PMM, is 14 kcal mol^{-1}, in excellent agreement with the value derived from experimental data. The protein environment is found to be responsible for lowering the energy barrier compared to the gas phase. A plausible PT pathway is hypothesized in which His51 is first protonated from the bulk solvent and then relays the proton via a bridging water molecule to the oxygen atom of Ser48 and subsequently to the substrate carbonyl. One or more water molecules forming a hydrogen-bond network between His51 and Ser48 are essential to activate the proton transfer chain.

The second study focuses on fatty acid photodecarboxylase from Chlorella variabilis (CvFAP), a recently discovered flavin adenine dinucleotide (FAD)-containing photoenzyme that catalyzes the light-driven decarboxylation of long-chain fatty acids into the corresponding alkanes, a reaction of considerable relevance for biofuel production. Despite its remarkable biotechnological potential, the mechanistic pathway of this trans-

formation remains only partially understood, encompassing uncertainties about the protonation state of the substrate, the origin of the unusual features in the enzyme's FAD absorption spectrum (which exhibits the main absorption peaks red-shifted by approximately 10–15 nm compared with those observed in other flavoproteins), and the nature of the red-shifted FAD species (FAD_{RS}) observed experimentally after decarboxylation.

Our QM/MM calculations strongly indicate that the bound fatty acid exists predominantly in its deprotonated form. Building on this result, the FAD absorption spectrum in CvFAP was simulated using MD–PMM calculations, which show that the contribution of flavin conformational bending to the observed red shift is minor (4 nm), whereas the protein environment is the dominant factor (15 nm). Regarding the origin of the FAD_{RS} species, our calculations support the hypothesis that its formation is likely driven by a pronounced reorganization of the active site following product formation (alkane, CO_2, and bicarbonate), with the bicarbonate interacting strongly with the FAD moiety.

Our results provide a solid basis for understanding the reaction mechanisms of these enzymes, offering mechanistic insights that could inform future environmentally friendly industrial applications. Moreover, from a computational point of view, this thesis demonstrates the complementarity of the two methodologies (QM/MM and MD–PMM) used here highlighting that their integration is fundamental for achieving a detailed and accurate characterization of enzymatic reaction mechanisms.

Contents

Abstract **1**

1 Introduction **6**

1.1 The importance of complex systems 6

1.2 Bio-inspired catalytic systems: the new frontier of industrial synthesis . 8

1.2.1 The role of dehydrogenases 11

1.2.2 Flavins and flavoproteins: an overview 13

1.2.3 CvFAP: the most promising photoenzyme for biocatalytic applications 15

1.2.4 Current challenges in biocatalysis 16

1.3 Hybrid computational methods: state of the art 17

1.4 Purpose of the thesis 22

2 Computational Methods **23**

2.1 QM/MM methods: state of the art 23

2.1.1 Additive QM/MM schemes 23

2.1.2 The ONIOM scheme 25

2.1.3 QM/MM: Limits and perspectives 30

2.2 MD-PMM: theoretical foundations 30

2.3 Computational details 34

2.3.1 *(S)*-Alcohol synthesis catalyzed by HLADH in the presence of a biomimetic cofactor 34

2.3.2 Mechanistic and spectroscopic study of Fatty Acid Photodecarboxylase . . . 37

3 Results and Discussion **41**

3.1 Modeling the HLADH-Catalyzed Formation of *(S)*-Alcohols . . . 42

3.1.1 Molecular Dynamics simulations . . . 47

3.1.2 Mechanistic investigation of the HT step . . . 49

3.1.3 Investigation of free energy barriers for HT step . . . 54

3.1.4 A possible pathway for the PT step . . . 57

3.1.5 Our mechanistic interpretation and future perspectives . . . 60

3.2 Mechanistic and spectroscopic study of fatty acid photodecarboxylase . 61

3.2.1 Active-Site Protonation state in Fatty Acid Photodecarboxylase 64

3.2.2 The simulation of FAD absorption spectrum in CvFAP . . . 68

4 Conclusions **85**

5 Appendix **90**

5.1 Simulation of Raman Spectra of Dicationic, Imidazolium-Based, Ionic Liquids . . . 90

5.1.1 Results and Discussion . . . 92

5.1.2 Conclusions . . . 100

5.1.3 Materials and method . . . 100

5.2 Supporting Information . . . 103

5.2.1 CvFAP . . . 103

5.3 List of publications . . . 110

5.4 List of abbreviations . . . 110

5.5 Acknowledgments . . . 112

List of Figures **113**

List of Tables **122**

Bibliography **124**

5.6 Copyright Note 143

Chapter 1

Introduction

1.1 The importance of complex systems

In recent years, there has been an increasing interest in the study of complex systems. This focus represents a direct response from the scientific community to recent global challenges such as the need to develop sustainable industrial processes in a period marked by climate change and resource scarcity.

A comprehensive understanding of complex systems requires moving beyond the simplistic description of these systems as mere aggregates of atoms. Instead, it is essential to conceptualize them as intricate networks of interacting components. These interactions give rise to emergent behaviors and properties that are not inherent to the individual isolated components [1]. In modern chemistry, complex systems, as summarized in figure 1.1, are generally large in size, show significant interactions among their components, may also display structural complexity, and can undergo substantial conformational changes over time. All of these features, taken together, give rise to an overall behavior that is nonlinear, hence it is not predictable considering system components as isolated. The biological realm represents the perfect example of complexity: nucleic acids, proteins, enzymes, photo-responsive systems, and biopolymers: all are characterized by an exceptional degree of complexity.

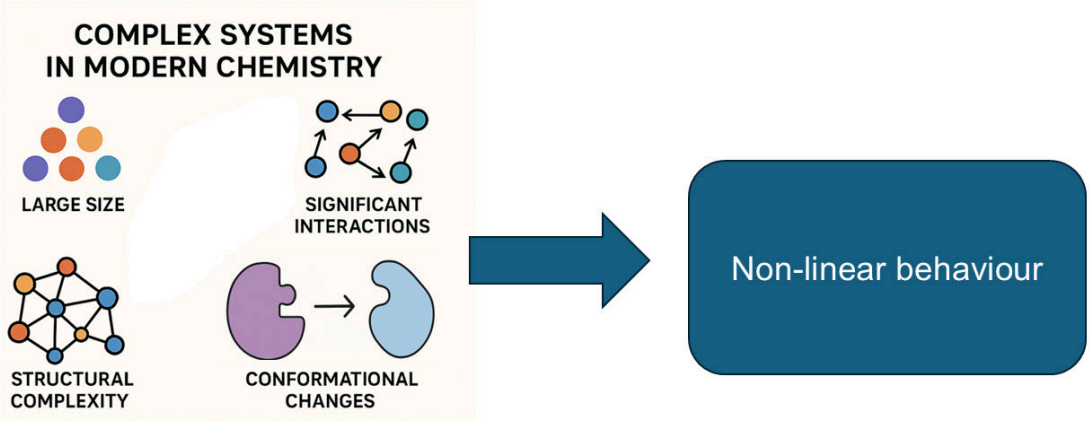


Figure 1.1: Key features of complex systems: large size, significant interactions between the components, structural complexity, and conformational changes. All these features result in a non-linear behaviour.

Among these, proteins represent particularly intricate macromolecular systems. They typically are large in size, frequently comprising thousands of atoms, and exhibit a multitude of significant interactions, not only between their constituent residues, but also involving specific residues and external entities such as the substrate and/or the solvent molecules. These interactions are often crucial, strongly influencing the chemical properties and the functional behavior of the protein. At the structural level, proteins display a hierarchical organization wherein each level (primary, secondary, tertiary, and quaternary) can influence the system's overall behavior [2]. Moreover, protein conformation is not fixed in time but undergoes substantial structural rearrangements, reflecting the intrinsically dynamic nature of these biomolecules.

To elucidate the dynamics of complex systems, researchers must perform multiple experiments that measure key variables under diverse conditions, thereby monitoring how the system responds to various perturbations over time. In this context, predictive models, combined with experimental data, assist them in formulating hypotheses that align with empirical observations and provide deeper insights into the behavior of complex systems [1]. It is therefore evident that the investigation and comprehension of biological mechanisms represent one of the biggest challenge currently faced by the modern theoretical chemists.

1.2 Bio-inspired catalytic systems: the new frontier of industrial synthesis

The origins of biocatalysis trace back over a century, when Rosenthaler performed the synthesis of (R)-mandelonitrile using a plant extract rich in hydroxynitrile lyase. This experiment was based on an innovative concept: using natural enzymes as viable alternatives to traditional synthetic catalysts in chemical processes [3].

Enzymes are, in fact, highly evolved biological catalysts that significantly accelerate chemical reactions within living systems. What makes biocatalysis particularly appealing nowadays is its alignment with green chemistry principles. Natural enzymes are more efficient than synthetic catalysts in terms of specificity and proficiency, and they typically work under mild conditions in non-toxic solvents. This not only enhances reaction yield but also significantly reduces environmental impact.

The exceptional catalytic efficiency and selectivity of enzymes are mainly related to their intricate interactions present in their extremely organized active site pockets where a network of strategically positioned residues orients the substrate with high precision thanks to several weak interactions such as: hydrogen bonds, electrostatic interactions, and Van der Waals forces [4].

Today, a vast array of enzymes has been identified, capable of catalyzing nearly all major classes of chemical reactions, including redox transformations, substitutions, additions, eliminations, rearrangements, and percyclic processes. Recent advances in elucidating enzymatic mechanisms have enabled the design of complex biocatalytic networks, where multiple enzymatic steps are integrated into synthetic pathways.

One of the most significant applications of biocatalysis is in asymmetric synthesis. This widely used technique enables the stereoselective production of enantiomerically pure compounds, from a prochiral precursor, and this is crucial in the pharmaceutical industry [5]. The pharmaceutical research, in fact, is the “engine” of advanced biocat-

alytic methods. This is attributable to the high value of pharmaceutical products, the strict purity requirements (especially concerning enantiomeric excess for chiral drugs), and the increasing pressure for more sustainable manufacturing practices.

A landmark achievement in industrial biocatalysis is the synthesis of sitagliptin, an antidiabetic drug. In its large-scale production, an engineered transaminase enzyme replaced a rhodium catalyst. This biocatalytic route led to significant improvements, including reduced costs, substantially lower waste generation, increased yields, and enhanced overall productivity [6, 4]. This successful result required the integration of both computational and experimental techniques.

Other applications of biocatalysis, as summarized in figure 1.2, can be found in the agricultural, chemical, cosmetic, and food industries. Recent examples include: the production of the synthetic polymer polyacrylamide, biofuels, plastic degradation, and the use of enzymes for the synthesis of pesticides. Additionally, the food industry makes use of biocatalysts for the production of additives (i.e., aspartame) and in baking processes.

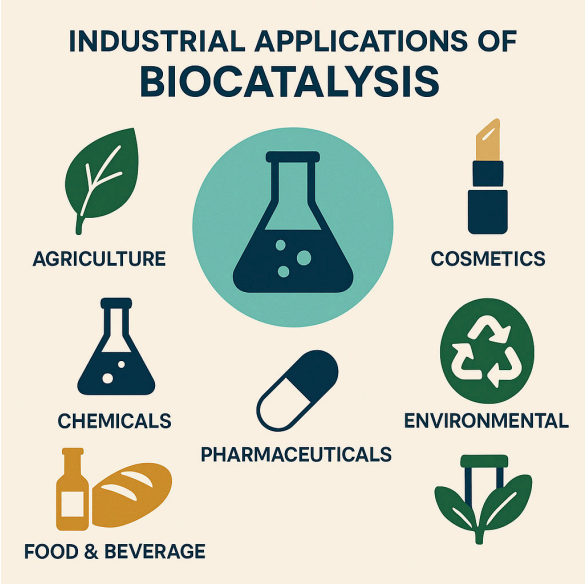


Figure 1.2: A summary of possible applications of biocatalysis.

In all these industrial processes redox reactions represent a central class of chemical transformations that are essential in a wide range of synthetic applications. In natu-

ral biological systems, these reactions are typically mediated by non protein organic molecules called cofactors, which facilitate the transfer of electrons with high specificity and efficiency. Among them, the most common are nicotinamide adenine dinucleotide (NAD^+) and flavin adenine dinucleotide (FAD).

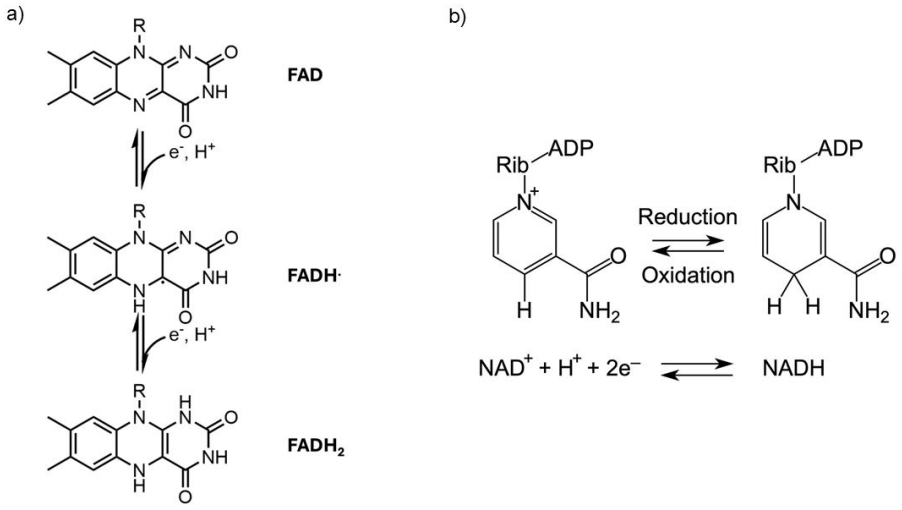


Figure 1.3: Redox states of FAD (a) and NAD^+ (b).

As illustrated in Figure 1.3, NAD^+ could easily be reduced to NADH accepting a hydride ion in a completely reversible redox reaction, generally coupled with substrate oxidation (or vice versa). Another important cofactor in biochemistry is $NADP^+$, which has the same structure of NAD^+ differing only by the presence of an additional phosphate group that replaces one hydroxyl group on the ribose subunit. FAD, a flavin-based cofactor, exhibits greater redox flexibility, adopting three distinct oxidation states [7, 8, 9].

Among the various enzymatic systems, this thesis focuses on two classes of enzymes of particular relevance to industrial biocatalysis: oxidoreductases, which catalyze redox (reduction–oxidation) reactions, and flavin-dependent enzymes, which are typically involved in photoinduced reactions of significant importance.

1.2.1 The role of dehydrogenases

Dehydrogenases are a prominent class of enzymes that catalyze a variety of biological redox reactions including: the Xylose oxidation, the oxidation of β-D-glucose to β-D-glucono-1,5-lactone, and the interconversion between L-glutamate to $\alpha-$ketoglutarate [10].

Alcohol dehydrogenase represents a particularly relevant subclass. These enzymes catalyze the reversible oxidation of alcohols to ketones (or the corresponding reduction).

The reduction reaction is generally more interesting due to their ability to preferentially produce one enantiomer (i.e. (S)-alcohol) starting from a prochiral substrate (i.e. ketone), makes them valuable tools for the synthesis of chiral alcohols that are key intermediates in the chemical and pharmaceutical processes. This redox reaction generally requires a cofactor (typically NADH or NADPH). From an industrial point of view, the cofactor is costly; hence, there is significant interest in developing processes that allow an easy and affordable regeneration of these cofactors [11, 12].

Among all oxidoreductases, horse liver alcohol dehydrogenase (HLADH) is a system of high relevance. Thanks to its broad substrate tolerance and its ability to efficiently catalyze the stereoselective reduction of ketones to *(S)*-alcohols with excellent stereoselectivity and high yields [13, 14], a key process for the pharmaceutical industry, HLADH has been the subject of extensive investigation by both experimental and computational research groups [15, 16, 17, 14].

From a structural point of view HLADH is a metalloenzyme that adopts a homodimeric quaternary structure, in which each monomer is organized into two polypeptide subunits: chain A and chain B. In the resulting dimer each chain contains two zinc ions(Zn^{2+}). The catalytically relevant zinc ion is tetrahedrally coordinated by Cys46, His67, Cys174, and a water molecule in the apo (without substrate) form of the enzyme; upon NADH cofactor binding, the water molecule is replaced by the ketone substrate (or the hydroxyl group in the case of the alcohol substrate) that directly

coordinates the Zn^{2+} ion maintaining the tetrahedral coordination. The second zinc ion is coordinated by four cysteine residues and has a structural role, not being involved directly in catalysis [11, 18]. The NADH cofactor has a key role in the reaction mechanism, not only because the reduction of the ketone is coupled to oxidation of the cofactor (NADH to NAD^+) but also the adenine dinucleotide moiety has a fundamental structural role, creating hydrogen bonds network that links the His51 residue with Ser48. This network plays a crucial role in the reaction mechanism. Further details will be provided in section 3.1.1.

The work of R. Fish et al. [14], surprisingly, demonstrated that the enzyme retains its catalytic activity even if NADH is replaced by smaller biomimetic cofactors lacking the adenine dinucleotide moiety, such as 1-benzyl-1,4-dihydronicotinamide (BNAH) (structure reported in figure 1.4). The reaction scheme proposed by Fish is extremely promising because the biomimetic cofactor has a reduced cost and could be easily regenerated at the end of the reactive cycle [19].

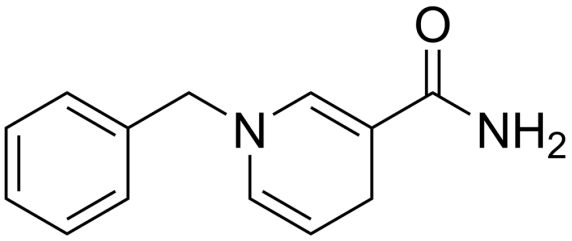


Figure 1.4: Structure of 1-benzyl-1,4-dihydronicotinamide (BNAH). In the work of Fish et al. [14] this molecule was used as biomimetic cofactor replacing the natural 1,4-NADH for the HLADH-catalyzed reduction of ketones to *(S)*-alcohol.

The notable study by R. Fish could pave the way for the practical implementation of HLADH in industrial applications; however, further research is required to gain a deeper understanding of the reaction mechanism when a biomimetic cofactor is employed.

1.2.2 Flavins and flavoproteins: an overview

Flavins are aromatic compounds derived from riboflavin (vitamin B_2) that have a key role in several biological processes such as redox reactions, signaling, DNA repair and circadian rythm regulation. [20, 21, 22, 23]. Enzymes that incorporate a flavin cofactor are called flavoproteins. The most common flavins present in biological systems are reported in figure 1.5. Their core isoalloxazine ring can adopt three redox states: oxidized, semiquinone, and hydroquinone, thereby enabling flavins to act as versatile cofactors in redox reactions [24, 25]. Although flavins vary structurally, they all preserve this ring, differing mainly in functional groups attached at the N_{10} position. Flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD), are the primary redox mediators in biological systems.

Flavins can be involved in ground-state redox reactions but, more importantly, can also undergo photoinduced reactions; hence they can act as chromophores in several photoreceptors and photoenzymes[26, 27, 28, 29]. Well-known examples are DNA photolyase and cryptochromes. DNA photolyases detect and repair DNA damage through a light-induced electron-transfer process whereas cryptochromes have a role in circadian rhythms regulation [30, 31]. Another key example is the riboflavin binding protein that is involved in the fastest electron transfer (ET) process occurring in flavoproteins [32]. From an industrial point of view, flavins are now being explored for biosensors, optogenetics, and photodynamic therapy [33, 34]. Accordingly, recently both experimental and computational groups have characterized many flavin proteins [35, 36, 37, 38, 39, 40].

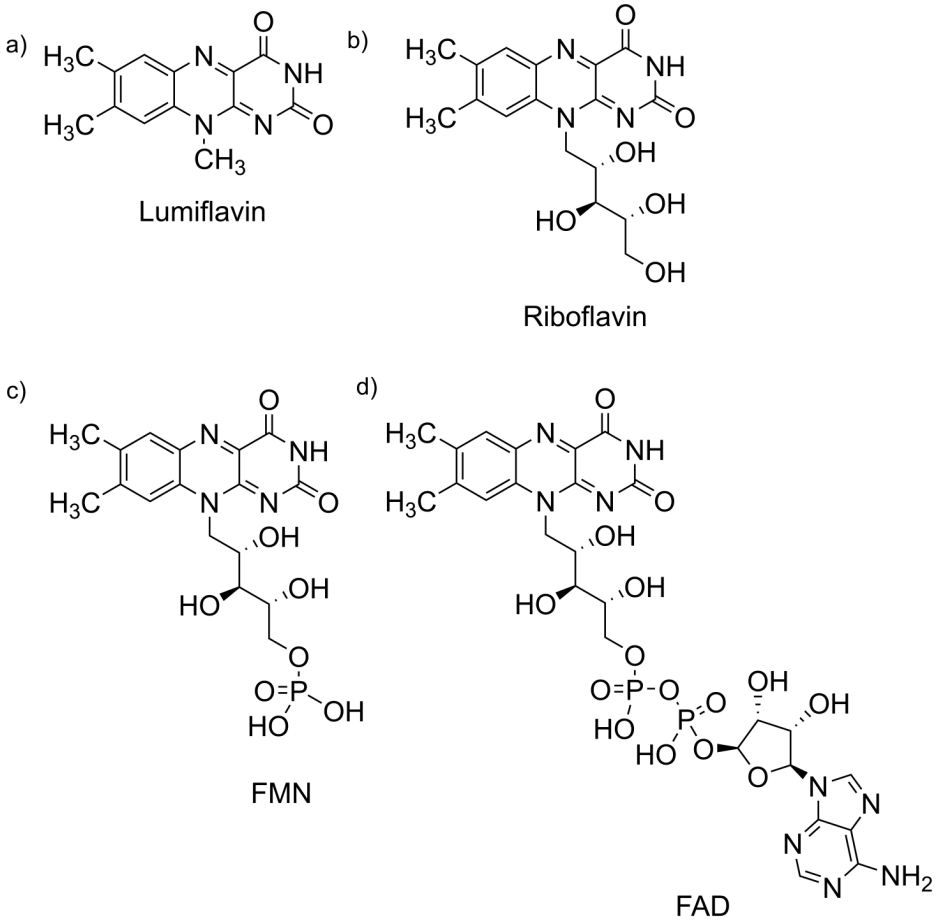


Figure 1.5: Structures of the most common flavins: a) lumiflavin, b) riboflavin c) FMN and d) FAD. Flavins are characterized by a common isoalloxazine ring, differing in the various functional substituents linked to that core.

From a computational standpoint, several studies were focused on lumiflavin, whose absorption spectrum serves as a starting point for the spectral simulation of more complex flavin cofactors such as FAD and FMN [22, 41]. Time-dependent density functional theory (TD-DFT) is generally the favored approach offering a good equilibrium between accuracy and computational cost. Kar et al. showed that the TD-DFT calculation conducted on lumiflavin in gas-phase at the B3LYP-ccpvDZ level of theory is in good agreement with data at higher level of theory (CC2 and ADC2), with an acceptable difference in bright peaks energies between CC2 and B3LYP (around 0.1 eV) [21]. However, in biological systems, the situation is more complex because the protein environment has a strong influence on the electronic structure thus strongly influencing the absorption spectra. To capture these effects, hybrid quantum-mechanics/molecular-

mechanics (QM/MM) schemes are widely used [42, 43, 22]. In FAD, the QM region is generally limited to the lumiflavin moiety, as its tail has a negligible contribution to the absorption spectrum as proved by several authors [21, 23, 44, 45]. Nevertheless, reproducing spectral shifts strongly influenced by protein conformational variations remains challenging.

For this purpose, a valid alternative to conventional QM/MM protocols could be the Molecular Dynamics–Perturbed Matrix Method (MD-PMM) that offers extensive conformational sampling while embedding environmental perturbations directly into quantum-mechanical observables [32, 40, 46, 47, 48, 49].

1.2.3 CvFAP: the most promising photoenzyme for biocatalytic applications

Recently, a newly discovered flavin system has attracted significant interest from the scientific community: the Fatty Acid Photodecarboxylase from Chlorella variabilis (CvFAP). This enzyme is the most recently discovered of only three known photoenzymes and incorporate FAD cofactor to catalyze the photoinduced decarboxylation of long-chain fatty acids to the corresponding alkanes with quantum yields exceeding 80% [50]. Additionally CvFAP could be also employed for the kinetic resolution of α-substituted acids to obtain chiral acids and more recently stereodivergently engineered variants of CvFAP were employed for the light driven synthesis of chiral alcohols and amines starting from the corresponding oxalates or oxamic acids with a chiral center at the γ-position [51]. This process is extremely promising since the first experiments showed a high yield, a wide substrate tolerance and an enantiomeric excess of around 99%.

Considering the potential of these applications in pharmaceutical and fine-chemical production, this enzyme represents one of the most promising systems in biocatalysis. However, its reaction mechanism is not yet fully understood, and further computational

and experimental studies are required to achieve a comprehensive elucidation of the reactive pathways, which is fundamental for future industrial development.

1.2.4 Current challenges in biocatalysis

Despite these significant advances, the industrial application of complex enzymes such as HLADH [11] remains challenging. These difficulties arise mainly from two factors: the incomplete elucidation of reaction pathways and the technical limitations associated with employing such enzymes in large-scale processes. Since enzymes are proteins evolved to work under cellular conditions, they could have a limited stability when exposed to non-physiological conditions typical of industrial settings. Elevated or reduced temperatures, extreme pH values, and atypical concentrations of substrates or products frequently lead to decreased activity or denaturation, limiting their practical applicability. To overcome these limitations, research funding toward biocatalysis and related disciplines is increasing in recent years. A central focus of these investments is the investigation of enzyme mechanisms, which represents the first step in the design of efficient biosynthetic pathways. In parallel, higher support for computational tools is accelerating the development of optimized biocatalysts [6].

From the perspective of computational chemistry, however, enzymes and biocatalytic systems present additional challenges due to their large size and structural complexity. Dealing with these systems often requires the adoption of hybrid computational methods, which combine different levels of theory to balance efficiency and precision. These methods will be discussed in the following section.

1.3 Hybrid computational methods: state of the art

Computational methods, with few exceptions, can be divided into two major categories: molecular dynamics (or sometimes, molecular mechanics) and *ab initio* calculations. Classical molecular dynamics (MD) [52, 53] simulates the time evolution of a molecular system integrating classical equations of motion. This approach, which often relies on empirical potential energy functions known as force fields, is a cornerstone of computational statistical mechanics. Thanks to its computational efficiency, MD allows the simulation of systems containing up to millions of atoms ($10^4 - 10^6$), making it particularly suitable for exploring phenomena at the border between chemistry and biology.

MD is capable of simulating process that occurs over relatively long timescales (from ns to ms) and to explore the conformational evolution of a system, but it is affected by few crucial limitations: in its standard form, it neglects electronic degrees of freedom, does not account for nuclear quantum effects, and relies heavily on the accuracy of the force field, which may fail to capture complex interactions.

In contrast, *ab initio* methods [54, 55, 56, 57] offer highly accurate descriptions of electronic properties by looking for approximate solutions of the Schrödinger equation from first principles. These methods are necessary for studying electronic structure, spectroscopy, and chemical reactivity. However, their applicability is restricted to relatively small systems (typically less than 500 atoms).

This creates a fundamental challenge in simulating complex chemical systems: while *ab initio* methods provide the necessary precision to capture electronic behavior, they lack the scalability to incorporate environmental effects that are essential for realistic modeling. The key to overcoming this apparent impasse lies in hybrid computational approaches that integrate the strengths of both paradigms. These multiscale or hy-

brid methods combine *ab initio* accuracy with the ability of molecular dynamics to deal with large structures, using a variety of schemes. Hybrid methods used in the study of biological systems are generally based on the QM/MM paradigm (Quantum Mechanics/Molecular Mechanics).

In the QM/MM formalism, as shown in fig. 1.6, the entire system is divided in two regions: an inner portion (denoted as quantum center or QM region) that is treated at electronic level and encompasses the atoms directly participating in the chemical process of interest, i.e. atoms involved in bond breaking or formation, a chromophore undergoing electronic excitation or a ligand and key active site residues in a protein-ligand complex; an outer layer, described by molecular mechanics, that comprises what is generally denoted as "the environment", i.e. the bulk of a protein, solvent molecules, or the extended framework of a materials [42].

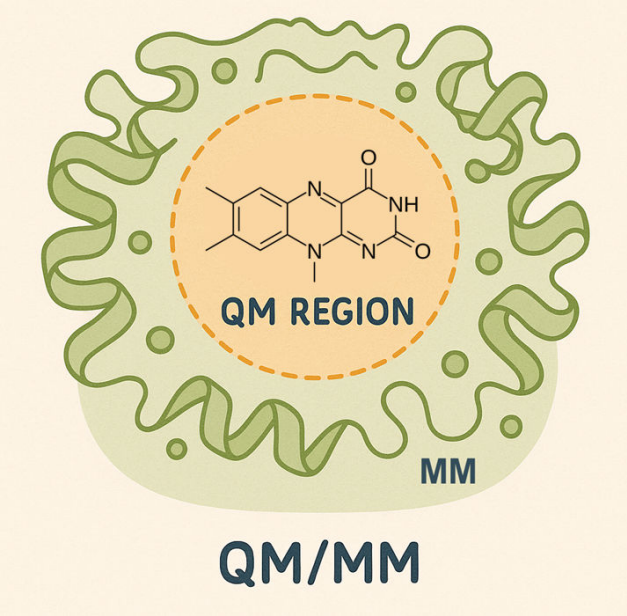


Figure 1.6: Basic QM/MM partitioning. The most chemically relevant part of the system is included (e.g. the lumiflavin moiety of FAD cofactor) in the QM part, while the MM part encompasses the remaining of the system (usually protein and solvent).

The choice of atoms to include in the QM region is crucial; it must be large enough to capture all essential quantum effects pertinent to the studied process but small enough to be treated with *ab initio* methods. The justification for this partitioning lies in the often-localized nature of chemical events, where the most significant electronic changes

are confined to a relatively small number of atoms, while the environment exerts its influence primarily through steric and electrostatic interactions. The selection of this QM/MM boundary is not a trivial task and can significantly influence the simulation accuracy and the computational cost [42, 58].

The QM/MM scheme is compatible with a wide range of combinations of different QM and MM methods. Generally in current biomolecular applications for the QM region, density functional method (DFT) is the best choice offering a good compromise between computational cost and accuracy. As an alternative, when the QM region is large or several calculations are required, semiempirical methods can also be employed. These methods are usually based on the use of minimal basis sets and involve strong approximations of the electron integrals. This approach leads to a significant reduction in computational time compared to traditional DFT methods [59]. Although this strategy entails reduced accuracy relative to full DFT, the recently developed tight-binding methods such as GFN2-xTB (Geometry, Frequency, Noncovalent — extended Tight Binding (version 2)) [57] represent a good compromise especially for geometry optimizations.

For mechanistic investigations in biomolecules, the initial stage of the calculation generally involves the screening of several possible reactive pathways. At this point, it is usually necessary to incorporate a considerable number of atoms in the QM region in order to examine their influence on the reactive state. For computational efficiency, the screening can be carried out using semiempirical methods to highlight the most probable reactive pathways among several possible candidates. Results can then be refined at a higher level of theory by considering only the most probable pathways. Following Grimme et al. [60], for the investigation of the HLADH reaction mechanism, we used this approach starting from the screening of six possible reactive pathways at the GFN2-xTB/MM level of theory, while the more demanding DFT-level free-energy calculations were performed on the two most likely candidate pathways.

The treatment of the MM region is generally achieved using classical MD where two

types of force fields are generally employed: additive all-atom force fields or polarizable force fields. The first category includes widely used force fields such as AMBER (Assisted Model Building with Energy Refinement) [61, 62], CHARMM36 (Chemistry at Harvard Macromolecular Mechanics) [63], or OPLS-AA (Optimized Potentials for Liquid Simulations-All Atoms) [64]. In this context, electrostatic interactions are commonly computed using an additive pairwise expression that employs fixed partial atomic charges. These force fields have been extensively employed for more than three decades. They are implemented in the most common molecular dynamics codes and have undergone numerous parametric refinements, generally yielding satisfactory accuracy in the simulation of complex biological systems. Nevertheless, such models are unable to account for charge redistribution in response to variations in the local electric field or in simpler words, they do not include induction energy and polarization effects [65, 66]. To address this limitation, polarizable force fields are sometimes employed. The most common polarizable force fields are AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Simulation) [67] (in its biomolecule variant) and CHARMM Drude [68]. In AMOEBA, the electrostatic energy is calculated by considering the interactions among atomic multipoles: each atom is in fact assigned a charge, a dipole, and a quadrupole moment. The electric field generated by these multipoles induces a dipole at each polarizable site, and the induced dipoles further polarize other sites in a self-consistent way [67]. Consequently, a realistic description of molecular polarization is achieved, albeit at a higher computational cost. Indeed, recent benchmarks show that classical non-polarizable force fields are typically 5–8× faster than AMOEBA [69].

In the Drude model adopted by CHARMM Drude [70, 71], an auxiliary mobile charged particle (the Drude particle) is attached to each atomic site by an harmonic spring. The atomic core charge q_{atm}, is adjusted such that the sum of the core and Drude charge q_d equals the original unperturbed static partial atomic charge q. The two point charges (q_{atm} and q_d) are separated by a variable displacement, $\mathbf{d}$ that changes

in response to the electric field($\mathbf{E}$), according to the expression

$$\mathbf{d} = \frac{q_D \mathbf{E}}{K_D}. \tag{1.1}$$

This expression allows to introduce the polarization of the atomic site where an induced dipole is now located and given by:

$$\boldsymbol{\mu} = q_D \, \mathbf{d} = \frac{q_D^2 \, \boldsymbol{E}}{K_D} \tag{1.2}$$

The computational cost of the CHARMM Drude force field (essentially due to the additional fictitious particles) is not excessively high, generally about twice that of the corresponding non polarizable force field, but it typically requires a time step of 1 fs, whereas the all-atom version usually employs a standard time step of 2 fs. Hence, the CHARMM Drude force field is generally four times slower than the corresponding non polarizable version

Another recently developed force field is GFN-FF (Geometry, Frequency, and Noncovalent Interaction Force-Field) . In this approach, electronegativity equilibrium (EEQ) model is used to calculate isotropic electrostatic energy and atomic partial charges. Here two sets of EEQ charges are used: one set is the standard, topology-defined charges, and the second is geometry-dependent. Thus, a partial polarization effect is introduced [72].

In the following chapter, further details about the QM/MM practical implementations will be discussed, with a particular focus on the schemes currently used to reproduce the polarization of the MM environment on the QM region, an aspect that is crucial for mechanistic investigations.

1.4 Purpose of the thesis

The main objective of this thesis is to elucidate the reaction mechanisms of two biocatalytic systems of high scientific and practical relevance: HLADH in the presence of a biomimetic cofactor, and the photoenzyme CvFAP. These studies were carried out using a combination of MD–PMM (Molecular Dynamics/Perturbed Matrix Method) [40, 73, 46] and QM/MM calculations [42]. The MD–PMM methodology will be described in detail in the following chapter.

By integrating these two approaches, combining accurate electronic-structure descriptions with extensive conformational sampling, to provide new mechanistic insights we will demonstrate the efficiency of these hybrid strategies for addressing the complexity of enzymatic reactivity and for guiding the rational design of more efficient, sustainable catalysts. Moreover the appendix contains a complementary study on the vibrational spectra of dicationic imidazolium-based ionic liquids, obtained through the integration of ab initio calculations and molecular dynamics, which confirms their theoretical reproducibility.

Chapter 2

Computational Methods

This section provides a detailed description of the methodological procedures employed in the studies presented in the Results section. In this thesis, the focus is placed on two main techniques, MD-PMM and QM/MM, for which the theoretical foundations are presented. By contrast, for the more established methods, including MD and Time-Dependent Density Functional Theory (TD-DFT), we report only the procedural details.

2.1 QM/MM methods: state of the art

2.1.1 Additive QM/MM schemes

In QM/MM calculations, two general approaches are commonly used to evaluate the total energy: additive and subtractive schemes [42, 58]. In the following discussion, the QM region will be denoted as QC, the MM region as E (Environment), and the entire system as S (System), where $S = E + QC$. The additive scheme is the more intuitive in formulation. Here, the total energy (E_{tot}) is expressed as

$$E_{tot} = E_{QM}(QC) + E_{MM}(E) + E_{Int}(QC - E) \tag{2.1}$$

Here $E_{QM}(QC)$ is the energy of the QM region calculated quantum mechanically as if it were isolated, $E_{MM}(E)$ is the energy of the MM region calculated at MM level and $E_{Int}(QC-E)$ is an interaction term which describes the coupling between the QM and MM regions [42, 74]. The last term includes bonded, van der Waals, and electrostatic interactions between QM and MM atoms. The electrostatic coupling between the QM and MM region could be described using various embedding models. In the mechanical embedding , the interaction between the QM and MM regions is completely handled at the MM level. Although this approach is straightforward and efficient, the QM density is not polarized by the environment, hence the accuracy is limited. A more sophisticated approach is the electrostatic embedding where the effect of MM charges is included in the QM Hamiltonian calculation as reported in equation 2.2 [58].

$$\hat{H}^{\text{total}}_{\text{QM-MM}} = \hat{H}^{0}_{\text{QM}} - \sum_{i \in \text{electrons}} \sum_{I}^{N} \frac{e^2 Q_I}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{R}_I|} \tag{2.2}$$

Here $\mathbf{r}_i$ and $\mathbf{R}_I$ are the positions of electron i and MM site I, $\hat{H}^{0}_{\text{QM}}$ represents the unperturbed one-electron operator for the kinetic and nuclear attraction energy of electron i and N represents the number of MM atoms that have a partial charge Q_I. As a result, the electrons perceive these MM atoms as sites bearing non-integer, and potentially negative, charges.

The main drawback of this model is that atomic charges in the MM region are typically fixed and non-polarizable. Despite this limitation, electrostatic embedding remains the most common approach in biological QM/MM applications and its accuracy is generally adequate for the majority of cases. In recent years, more advanced strategies, denoted as polarizable embedding models have been developed. These approaches allow mutual polarization between the MM environment and the QM electron density, typically through a complex iterative algorithm. Ganguli et al. [75] provided an extensive benchmark on enzymatic reactions, comparing QM/MM free energy barriers obtained with polarizable embedding using the CHARMM Drude force field to those computed with the conventional electrostatic embedding approach based on the CHARMM22

force field. Their results showed that explicit MM polarization only moderately affects (2-3 kcal mol^{-1}) the calculated activation free energies, while significantly increasing the computational cost (7-8 times slower).

Nevertheless, in cases involving large changes in electrostatic properties or that require very high accuracy the explicit inclusion of polarization becomes essential. It is therefore important to carefully identify the systems where the polarization embedding is necessary. Another crucial factor is the conformational sampling of the protein environment. Recent studies have demonstrated that, with sufficient sampling, even fixed-charge force fields can reproduce complex properties such as pK_a values and reduction potentials. By contrast, considering only a single representative structures can lead to large average errors [76].

2.1.2 The ONIOM scheme

Alternatively to the additive schemes, another approach is the subtractive model where the total energy is given as:

$$E_{total} = E_{QM}(QC) + E_{MM}(S) - E_{MM}(QC) \tag{2.3}$$

Its accuracy relies on the assumption that the MM force field can provide a reasonable description of the QM region and that errors arising from different terms tend to cancel. An important example of a subtractive scheme is the ONIOM method (Our Own N-layered Integrated molecular Orbital and molecular Mechanics). Originally developed in 1995 by Morokuma and colleagues as the "Integrated Molecular Orbital + Molecular Mechanics (IMOMM)" method [77, 78], ONIOM, in its refined version, is nowadays considered the most popular approach for QM/MM simulations, particularly for biological systems.

This scheme is simpler than the classical additive scheme, because there is no additional QM–MM coupling term in the Hamiltonian. Instead, an MM calculation of the QM

region is required. However in the standard ONIOM energy expression, the mutual interactions between QM atoms are included, at MM level, both in $E_{\mathrm{MM}}(S)$ and in $E_{\mathrm{MM}}(QC)$ terms and therefore this contribution is reciprocally canceled in the final energy expression.

In the original ONIOM formulation the electrostatic embedding is absent; this means that the QM region electrostatic is not influenced by the MM region. The interaction between QM and MM layers is described only by the $E_{\mathrm{MM}}(S)$ term at the MM level [43].

To improve the accuracy of this scheme, an "ONIOM electrostatic embedding model" has recently been developed [79]. It should be noted that this approach differs from the electrostatic embedding employed in additive QM/MM schemes. In ONIOM electrostatic embedding, the charges from the MM region are incorporated into the Hamiltonian of the QM calculation $E_{\mathrm{QM}}(QC)$, and in the $E_{\mathrm{MM}}(QC)$ term they are treated at the classical level, keeping the $E_{MM}(S)$ term unchanged from the one already reported in the ONIOM mechanical-embedding energy expression.

The two-layer ONIOM scheme can be extended to a three-layer model that includes an intermediate layer (QM2) between the high-level quantum mechanics (QM1) and molecular mechanics (MM) regions. This approach is particularly useful for studying complex chemical reactions in systems with large reactive centers ($>$ 100 atoms). In this framework, the QM1 region is treated with high-level methods such as post-Hartree-Fock methods or Density Functional Theory (DFT), while the intermediate QM2 region is described by a lower-level theory, such as semiempirical methods or less demanding DFT functionals. The remaining part of the system is represented by a classical MM force field. The total energy for a three-layer system is given by the expression:

$$E_{\mathrm{ONIOM3}} = E_{\mathrm{high},QM1} + E_{\mathrm{medium},I} - E_{\mathrm{medium},QM1} + E_{MM,S} - E_{MM,I} \qquad (2.4)$$

Herein, S designates the total system, and I denotes the intermediate system, defined as (I=QM2+QM1).

The methodology employs two levels of *ab initio* theory: a "High" level of greater accuracy and a "Medium" level of lower accuracy. The "High" level is applied to the QM1 region, whereas the 'Medium' level is employed for the QM2 region.

Figure 2.1 summarizes the ONIOM three layers scheme.

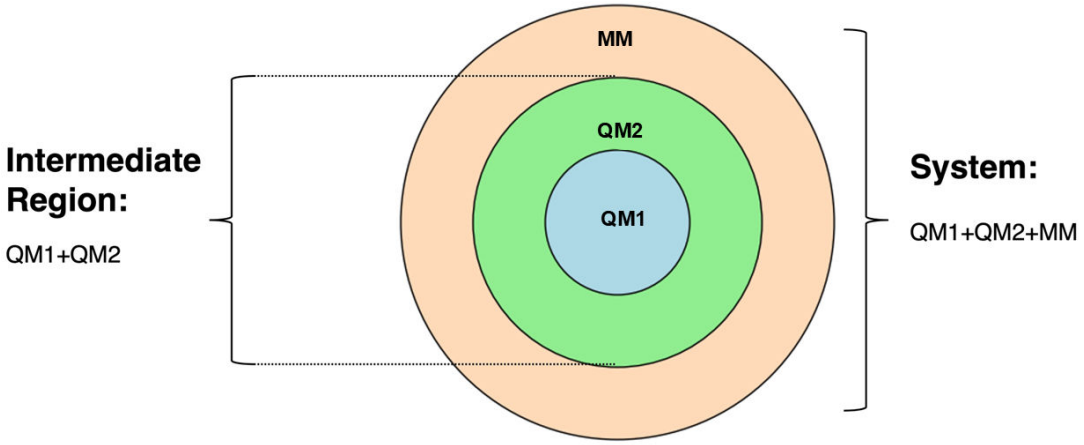


Figure 2.1: ONIOM Three Layers system division The QM1 is the region where the process of interest occurs, generally modelled with high level DFT. The QM2 region is the "surrounding" of the QM1 region, generally treated at semiempirical level, while the MM region is the external environment.

A significant, inherent limitation present in the ONIOM method and in all QM/MM approaches, is the necessity of partitioning the system into distinct QM and MM regions. In biological systems, this division could be particularly challenging. The boundary between the QM and MM regions, or between the high-level QM1 and low-level QM2 layers in a multi-layered ONIOM scheme, often requires the cleavage of a covalent bond. This issue is typically addressed through the link-atom approach [43], where the valency of the broken atomic fragments is saturated, most commonly with a hydrogen atom.

The established practice in biochemistry is cutting the single, nonpolar C_α-C_β bond in amino acids. To better understand this concept, it is possible to consider the case of a cysteine residue. Here the fragment treated at QM level becomes a methylthiol model

system (CH_3-S), while the remainder of the residue is handled at the MM level. In more intricate molecular situations, however, the selection of an appropriate partition can be not trivial. While general guidelines have been proposed in the literature [80], such as avoiding the cleavage of double, peptidic or strongly polarized bonds, the decision remains substantially arbitrary and constitutes a notable weakness of the methodology. In particular, the inclusion of backbone segments in the QM2 region (or in the QM region in two layers schemes) can create artifacts in the QM layer, with a potential impact on the system's stability and, in some cases, leading to unreliable results.

A critical preliminary step in the investigation of biological systems through hybrid computational methodologies is the optimization of the initial structural configuration. The selection of the optimization algorithm is crucial, as it can significantly influence the accuracy and reliability of the results and also the computational cost.

Considering the classical partitioning adopted in QM/MM, it is evident that the environment possesses a greater number of degrees of freedom. However, the computational cost associated with molecular mechanics (MM) calculations for the full system is significantly lower than that of quantum mechanical (QM) calculations for the inner region.

The procedure commonly implemented for geometry optimizations within the ONIOM framework is typically based on the "micro-iterations" approach, summarized in figure 2.2. Initially, a full MM geometry optimization of the environment region is performed keeping the QM region fixed. Subsequently, the most computationally demanding step is carried out: the QM optimization of the inner region, with the environment held fixed. This cycle is repeated iteratively until convergence criteria are met.

This two-step optimization approach not only reduces the number of expensive QM iterations but also allows the use of distinct, efficient optimization algorithms specific for each region (RFO for the inner region and Steepest descent for the MM region are the most common). As a result, the overall computational efficiency of the method is

significantly enhanced.

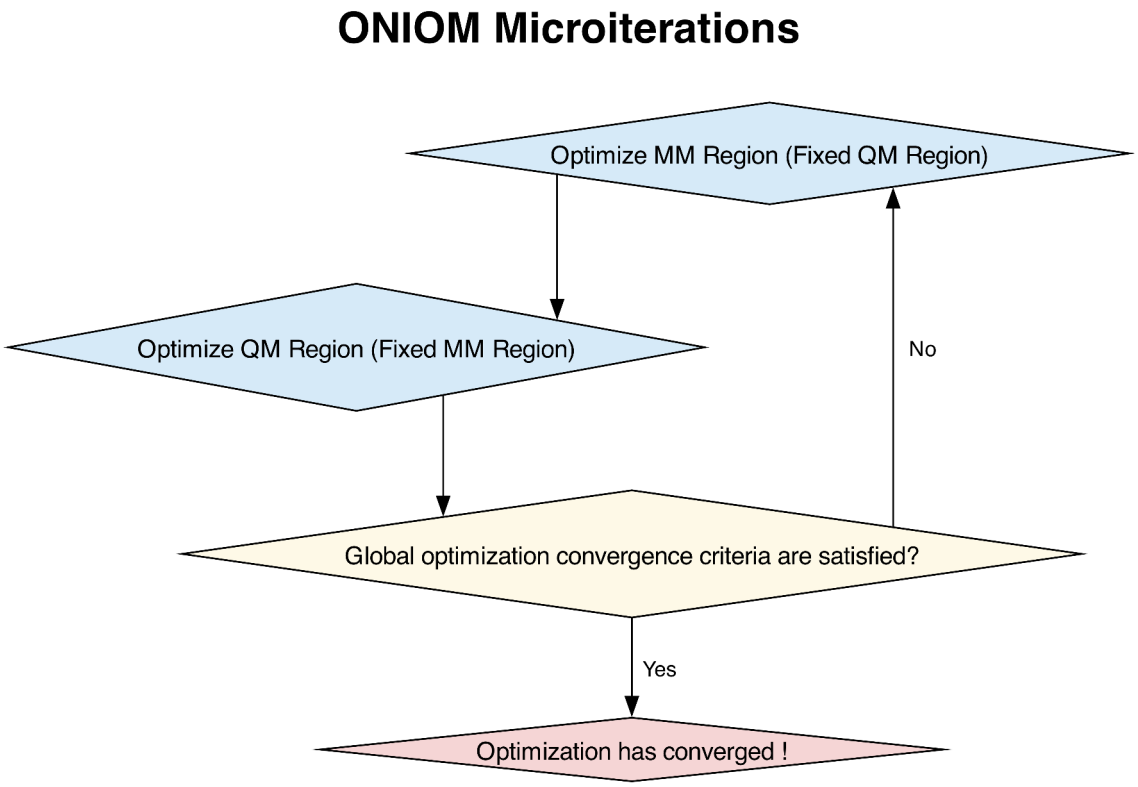


Figure 2.2: ONIOM Microiteration algorithm

The drawback of this approach is that it excludes Hessian coupling between the QM coordinates and MM coordinates, resulting in possible oscillations in the optimization steps. Alternative approaches are based on the quadratically coupled optimization method that allows explicit Hessian coupling between the QM and MM parts in the QM optimization steps. This approach makes the optimization more rigorous improving also the stability but it has higher computational cost compared to microiterations algorithm.

This more evolved approach is implemented in Gaussian software, however the microiterations algorithm is usually preferred due to its simplicity.

2.1.3 QM/MM: Limits and perspectives

QM/MM methodologies possess several inherent limitations. The first is the arbitrary partitioning of the system into the QM and the MM regions. This division sometimes makes the treatment of the QM-MM boundary problematic because the truncation of covalent bonds could introduce significant artifacts. Furthermore, the polarization of the QM region is often inadequately described, even within electrostatic embedding schemes. These models typically rely on fixed MM point charges, which are inherently static and can lead to inaccuracies. In contrast, fully polarizable embedding models are accurate but computationally expensive and are therefore generally recommended only for specific cases.

A more critical limitation of conventional QM/MM methods is their inherently static nature. They rely on the energy minimization of a single or a few representative structures, neglecting the dynamical behavior and conformational mobility of the systems. In biological systems, however, capturing temporal evolution and structural fluctuations is often essential for an accurate and realistic description of spectroscopy and reaction pathways. For this specific purpose, we introduce a method that accurately accounts for the perturbation exerted by the environment on the QM region, while also considering an extensive conformational sampling of the protein: the Molecular Dynamics/Perturbed Matrix Method (MD-PMM) [73, 46].

2.2 MD-PMM: theoretical foundations

The MD-PMM method, developed by Amadei et al. [73, 46], enables the inclusion of dynamical effects at reduced computational cost. Also in the MD-PMM approach the system is divided in the QM region (denoted as QC) where the process of interest occurs and in the MM region that represent the environment. The theoretical foundation of the MD-PMM method lies in perturbation theory, wherein the Hamiltonian of the QC

can be expressed as:

$$\widehat{H} = \widehat{H_0} + \widehat{V} \tag{2.5}$$

Here $\widehat{H}$ represent the perturbed Hamiltonian, while $\widehat{H_0}$ represents the unperturbed Hamiltonian and $\widehat{V}$ is the perturbation operator.

In this work, the dipole-based MD-PMM variant was employed, as it had already been successfully applied to simulate the absorption spectrum of riboflavin in aqueous solution [40]. The previously reported results were highly satisfactory: the dipole-based MD-PMM accurately reproduced the peak positions, with a mean absolute error (MAE) of only 6 nm, and the relative intensities of the two riboflavin peaks were in excellent agreement with the experimental data. Moreover, the same approach was able to describe the ultrafast photoinduced electron-transfer (ET) kinetics between riboflavin and riboflavin-binding protein in the excited state, yielding results consistent with experimental observations [40]. Here we report the essential theoretical aspects.

The typical MD-PMM procedure could be divided in different steps. The initial phase involves an extensive MD simulation to sample the conformational ensemble of the system. From this trajectory, one or more (depending on the flexibility of the QM region) representative structures of the QC are extracted. In traditional MD-PMM applications [40, 81], the quantum mechanical calculations on the QC structures are generally performed in gas-phase after a fast optimization procedure (generally with very loose convergence criteria).

Here, as reported in figure 2.3, we present a different approach in which an additional and optional intermediate QM/MM step, starting from a structure of the whole system extracted by the MD simulation, is performed prior to the calculation of the unperturbed Hamiltonian of the QC (see section 3.14 for more details). The unperturbed Hamiltonian was therefore computed on the gas-phase QC structure, which was extracted from the QM/MM-optimized structure and subsequently reoptimized in the gas phase by applying strong constraints to selected degrees of freedom, usually to all heavy atoms. This ensures that the geometry of the reaction center remains consistent

with that already optimized within the protein environment. The perturbed Hamiltonian matrix is then constructed and diagonalized at each frame of the MD trajectory, according to the expression reported in equation 2.6:

$$\widehat{H} = \langle \psi_n^0 | \hat{H} | \psi_{n'}^0 \rangle = \widehat{H_0} + I q_{tot} V - \mathbf{E} \cdot \langle \psi_n^0 | \hat{\boldsymbol{\mu}} | \psi_{n'}^0 \rangle \tag{2.6}$$

V and $\mathbf{E}$ represent the electrostatic potential and electric field induced by the environment on the QC; q_{tot} is the total charge; ψ_n^0 corresponds to the unperturbed (gas-phase) electronic eigenstates of the QC; I stands for the identity matrix; and $\psi_{n'}^0$ refers to the n'-th associated excited state.

From the diagonalization of this matrix at each MD frame, the QC's properties and observables of interest, such as the perturbed energy, can be evaluated as a function of time. In some cases, the same procedure is repeated for different molecular dynamics trajectories (or for different segments of the same trajectory), and the average value of the variable of interest is then obtained.

Through this computationally efficient and straightforward procedure, MD-PMM is capable not only of incorporating the effect of the environmental perturbation on the QC but also of accounting for the conformational changes of the system over time.

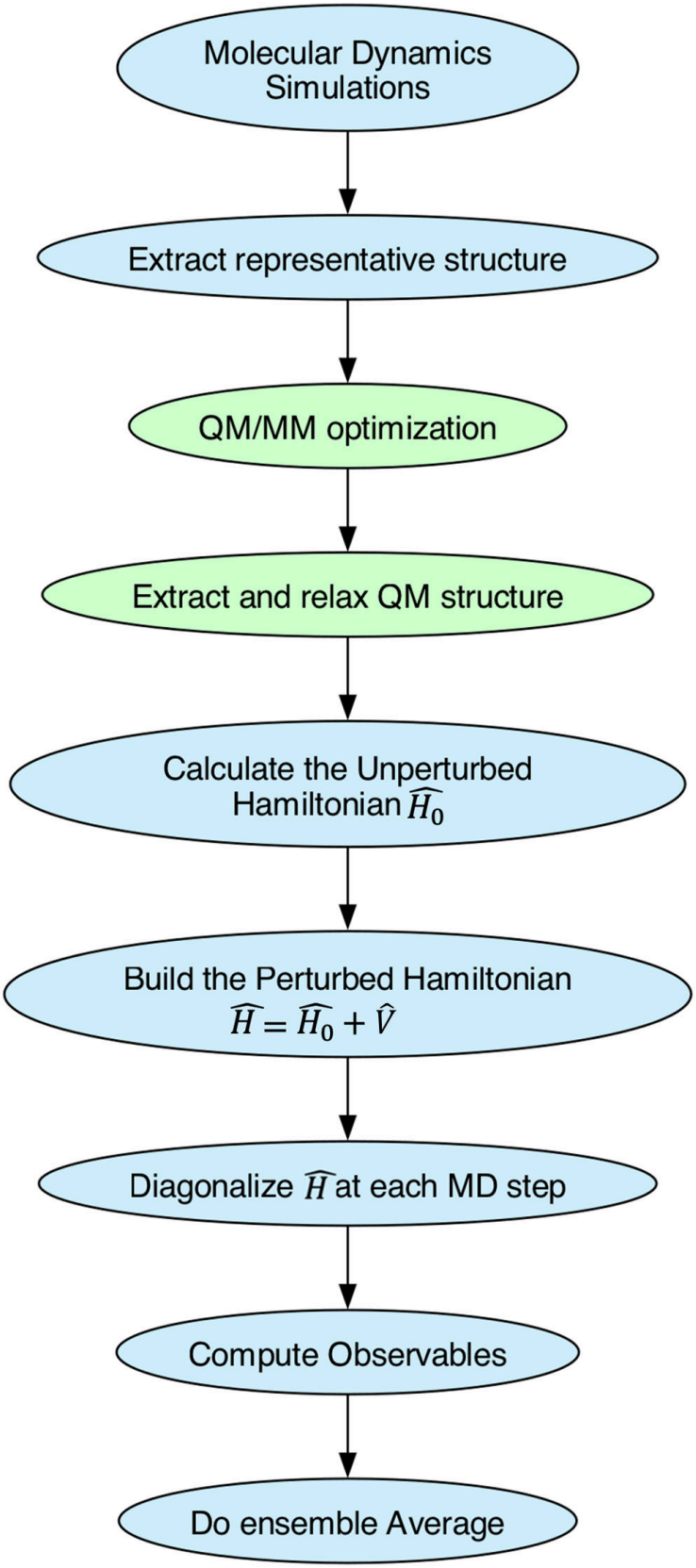


Figure 2.3: Workflow of a typical MD-PMM calculation The key step is the diagonalization of the perturbed Hamiltonian matrix, including the effect of the environmental perturbation at each MD frame. In green we have highlighted the QM/MM step introduced in this work.

2.3 Computational details

2.3.1 *(S)*-Alcohol synthesis catalyzed by HLADH in the presence of a biomimetic cofactor

MD simulations

The initial coordinates for the MD simulations were taken from the refined crystal structure of the HLADH (PDB ID: 7K35). A 2-pentanone molecule was inserted in the active site by aligning it with the co-crystallized substrate, 4-methylbenzyl alcohol. Additionally, the N-Benzyl-1,4-dihydronicotinamide biomimetic cofactor replaced the co-crystallized 1,4-NADH cofactor building on previous data reported by Marrone et al. [16, 19]. The protonation states of histidine residues were determined using the PropKa server [82]. For His51, two distinct charge states were considered: a neutral state with the proton located on the N_ε atom (referred to as HIE51), and a cationic state (+1), in which both N_δ and N_ε are protonated, denoted as HIP51.

All MD simulations were performed using the GROMACS [83] software package. The protein was modelled with the AMBER99 [61] force field, while the GAFF (General Amber Force Field) [62] force field was employed for the organic ligands (2-pentanone and N-Benzyl-1,4-dihydronicotinamide). The active site was modelled using a bonded model adapted from Lin et al. [84], wherein the Zn^{2+} is coordinated by the sulfur atoms of negatively charged (-1) Cys46 and Cys174 residues, the oxygen atom of the substrate, and the N_ε atom of His67 (neutral). The complete protein-ligand system ($\approx$ 11350 atoms) was placed in a cubic box (in periodic conditions, box size=13.6 nm), solvated with 79450 TIP3P [85] water molecules, and neutralized with an appropriate number of counterions. Long-range electrostatic interactions were handled with the Particle Mesh Ewald (PME) method [86] with a real-space cutoff of 1.1 nm. The Lennard-Jones potential cutoff was also set to 1.1 nm. The time step was set as 1 fs

in the production run and 2 fs for the equilibration part. The LINCS algorithm was used to constrain bonds involving hydrogen atoms [87].

The simulation protocol consists of an initial energy minimization, followed by a 100 ps annealing phase where the temperature was increased from 50 K to 300 K. Subsequently, a 2 ns equilibration was performed in the NPT ensemble using the V-Rescale thermostat (τ = 0.002 ps, T = 300 K) [88] and the Berendsen barostat (τ = 1 ps, P = 1 bar) [89]. During these initial steps, all protein heavy atoms were restrained with a force constant of 1000 kJ mol^{-1} nm^{-2}. These restraints were then removed, except for the atoms of the reaction center (Cys46, Ser48, His51, His67, Cys174, Zn^{2+}, and both ligands). The system was then subjected to 3 further equilibration runs with progressively reducing the restraint constant from 1000 to 100, and finally 10 kJ mol^{-1} nm^{-2}. After equilibration the unrestrained production run was conducted for 50 ns using the V-Rescale thermostat (τ = 0.002 ps, T = 300 K) [88] and the Berendsen barostat (τ = 1 ps, P = 1 bar) [89].

The analysis of hydrogen bond networks involving water molecules between His51 and Ser48 residues was conducted using an in-house tool, according to hydrogen bond criteria by Kumar et al. [90].

ONIOM QM/MM optimizations

To explore all possible reaction pathways, we carried out potential energy scans (PES) within a QM/MM framework. Geometry optimizations were conducted with the ONIOM method [43, 77] implemented in the xTB program [60]. The QM layer was treated at the GFN2-xTB level [57] and included Cys46, Ser48, His51, His67, Cys174, the Zn^{2+} ion, the 2-pentanone substrate, the biomimetic cofactor, and one water molecule. The remainder of the system was described by the GFN-FF force field [72].

The multiscale GFN2:GFN-FF optimizations were performed with the FIRE algorithm [91] in generalized Born implicit solvent [92] until convergence thresholds of

0.002 hartree/bohr (maximum gradient) and 0.00005 hartree (energy change) were satisfied.

Three representative optimized structures were then used as starting points for constrained optimizations along selected reaction coordinates (PES scan), yielding potential energy profiles for all candidate pathways (Model I-VI, see the Results chapter for details). Apart from the reaction coordinate, no additional geometric constraints were imposed.

MD-PMM calculations

The calculation of the Gibbs free energy barrier for the hydride transfer step for the two selected reaction mechanisms (denoted as I and V, see the Results chapter for details) was conducted by means of the MD-PMM method. Three representative geometries, corresponding to the reactant, the highest-energy point (i.e., the geometry closest to the transition state), and the product, were extracted from the corresponding ONIOM QM/MM potential energy scans. For mechanism V, a fourth geometry corresponding to the η^2-complex was also included. As the QC reference, the same QM region used in the QM/MM calculations was adopted. The Hamiltonian of each isolated QC, $\hat{H}_0$, was evaluated at the B3LYP/6-31G* [93, 94] level of theory in gas-phase. At each MD step (see the Results chapter for the details of these MD simulations), the perturbed Hamiltonian $\hat{H}$ (see Eq. 2.6), was subsequently calculated and diagonalized to obtain the corresponding perturbed electronic energies (E_i).

The standard Gibbs free–energy change between two given points of the reaction pathways was then computed according to Eq. 2.7:

$$\Delta G^{\circ} = -\, k_{\mathrm{B}} T \ln \left[\frac{1}{N} \sum_{i=1}^{N} \exp\left(-\frac{\Delta E_i}{k_{\mathrm{B}} T} \right) \right], \tag{2.7}$$

where k_{B} is the Boltzmann constant, T the absolute temperature, N the number of MD frames, and ΔE_i the perturbed electronic energy difference between the two states

at each frame i.

2.3.2 Mechanistic and spectroscopic study of Fatty Acid Photodecarboxylase

MD simulations

All MD trajectories were conducted with the GROMACS software package [83] and the CHARMM36 force field [95, 96]. The apo configuration was prepared starting from the crystal structure of CvFAP (PDB ID: 6YRU) [97] by removing the substrate from the active site.

The *apo* system ($\approx$ 8600 atoms) was inserted in a periodic dodecahedral cell ensuring a minimum distance of 1.2 nm between the protein and the edges of the simulation box and solvated with 23000 TIP3P water molecules [85]. Charge neutrality was ensured by adding an appropriate number of Na^+ counterions. Long-range Coulomb interactions were treated with the Particle Mesh Ewald (PME) scheme [86], using 1.2 nm real-space cutoff. The same cutoff was also used for Lennard–Jones interactions. Bond lengths involving hydrogen atoms were constrained using the LINCS algorithm [87].

During energy minimization, annealing, and equilibration stages, harmonic positional restraints (force constant $k = 1000\ \mathrm{kJ\,mol^{-1}\,nm^{-2}}$) were applied to protein C_α atoms and to the heavy atoms of the FAD cofactor. During annealing the temperature was raised from 50 K to 300 K over 100 ps. This step was followed by a 2 ns equilibration in the NVT ensemble at 300 K using the Berendsen thermostat (relaxation time $\tau = 0.5$ ps), and subsequently by 1 ns of NPT equilibration at 300 K and 1 bar also employing the Berendsen thermostat and barostat (relaxation time $\tau = 0.5$ ps; compressibility $= 4.5 \times 10^{-5}\ \mathrm{bar^{-1}}$) [89]. A time step of 1 fs was used for the equilibration stages.

Production runs consisted of a 100 ns trajectory in the NPT ensemble at 300 K and

1 bar, with a 2 fs integration step. Temperature is controlled with the Nosé–Hoover thermostat (relaxation time 1.0 ps) [98], while for pressure control the Parrinello–Rahman barostat (relaxation time 5.0 ps; compressibility = 4.5×10^{-5} bar^{-1}) [99] was employed.

QM/MM optimizations

The initial structure for the 300 K simulations was based on the refined crystal structure of the *Holo* CvFAP at pH 8.5; in this model, stearic acid was replaced by palmitic acid. The protonation states of all ionizable residues were assigned using PropKa [82]. For the substrate, two protonation forms were considered: palmitic acid (neutral) and palmitate (anionic). The energy minimization was performed using the AMBER force field[62, 61] and the ORCA 6 MM module[100, 101], with all heavy atoms constrained. A subsequent QM/MM optimization was conducted with the electrostatic embedding scheme [102, 42, 103]. At the QM/MM boundary, a charge-shifting scheme was applied, and the bonds were saturated using hydrogen link atoms[42, 80, 104]. For the MM part the AMBER force field was used to retain compatibility with the Orca code.

The QM region included the Arg451 side chain, palmitate head group (-CH_2-CH_2-CO_2^- fragment), lumiflavin moiety of the FAD cofactor, and two water molecules hydrogen-bonded to palmitate (Wat1 and Wat2) and, in some calculations, also the Cys432 side chain was included in the QM region. For the product structure, the head of palmitate and Wat1 were replaced by a bicarbonate ion and the head of pentadecane (-CH_2-CH_3 fragment). The QM region was modelled with DFT (B3LYP or CAM-B3LYP) functionals using the def2-TZVP basis set [93, 94, 105] . The RIJCOSX approximation [106], combined with its auxiliary basis, was used to improve computational efficiency. The MM portion of the system, modeled with the AMBER force field, was divided into an active mobile shell and a static part. Around the QM core, 350 atoms formed the active MM region, including Ile130, Phe134, Ala171, Leu173, Cys432, Arg451, Val463, Tyr466, Gln486, Ser573, Ser574, Asn575, Gly622, and the remaining fragments of FAD and palmitate. All remaining atoms are included in the static MM region.

Geometry optimizations were performed using the BFGS method [107] adopting the default convergence criteria implemented in ORCA 6 [108] package. Only crystallographic water molecules were included in the calculation.

From the optimized structures, TD-DFT UV–Vis spectra of the reactant and product were computed in the same QM/MM framework without the TDA approximation [109], at the B3LYP/def2-TZVP (or CAM-B3LYP/def2-TZVP) level, including the first twelve singlet excited states. For the assessment of the palmitic acid protonation state (RCOOH/ $RCOO^-$), the TD-DFT calculations were repeated including the first 25 singlet states to find the high-energy CT states correlated with each configuration.

MD-PMM calculations

Three representative geometries were extracted from the apoprotein molecular dynamics simulation considering three different degrees of lumiflavin ring bending. To retain the structural conformations, six dihedral angles involved in the "butterfly bending" motion were constrained during geometry optimizations in the gas phase at the B3LYP/def2-TZVP level of theory. The constrained dihedrals were : C_4–N_5–N_{10}–C_9, C_6–N_5–N_{10}–N_1, C_7–C_6–C_{5a}–N_5, C_8–C_9–C_{9a}–N_{10}, N_{10}–C_{10a}–N_1–C_2, and N_5–C_{4a}–C_4–N_3. Atoms numbers are reported in fig. 2.4

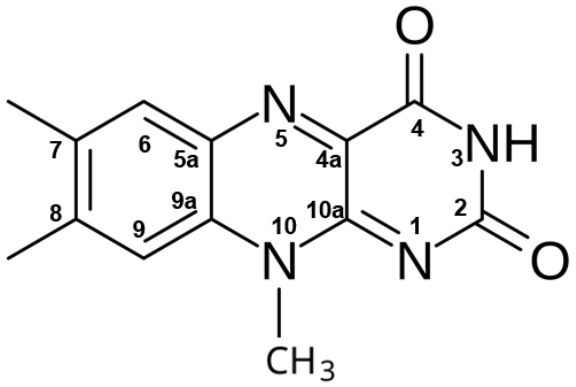


Figure 2.4: Atom numbers assigned to the lumiflavin moiety of the FAD cofactor

On these optimized geometries, TD-DFT calculations were performed to compute the lowest twelve singlet excited states, using the same level of theory and without the

Tamm-Dancoff approximation TDA. In the MD-PMM calculation only the first ten excited states were considered.

TD-DFT calculations provided the unperturbed ground-state energy and the excitation energies of the first ten excited states. Additionally dipole moments associated to both ground and excited states were also calculated. Transition dipole moments between excited states were omitted, consistent with prior studies [110, 40] which have shown their contribution to the absorption spectrum to be negligible. At each frame of the MD trajectory, the perturbed Hamiltonian matrix was constructed and diagonalized according to Equation 2.6, yielding the instantaneous perturbed transition frequencies (ν) and transition dipole moments (μ_{ij}). To calculate the absorption spectrum, the distributions of transition frequencies and dipole moments were evaluated by discretizing the frequency space into finite bins. These bins were subsequently employed to determine the molar extinction coefficients, $\varepsilon_{0,n}$, associated with ground-to-nth excited state transitions. The absorption spectrum was then obtained using standard expressions. The spectral line shape was calculated according to the following equations 2.8 and 2.9:

$$\varepsilon_{0,n}(\nu) = \sum_{\nu_{\text{ref}}} \frac{T_A(\nu_{\text{ref}})\eta(\nu_{\text{ref}})h\nu}{N} \cdot \frac{e^{-\frac{(\nu-\nu_{\text{ref}})^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \tag{2.8}$$

$$T_A(\nu_{\text{ref}}) = \frac{|\mu_{0,n}|^2_{\nu_{\text{ref}}}}{6\varepsilon_0 c\hbar^2} \tag{2.9}$$

Here, ν represents the central frequency of each interval, $\eta(\nu_{\text{ref}})$ denotes the number of MD frames contributing to that frequency bin, and $|\mu_{0,n}|^2_{\nu_{\text{ref}}}$ corresponds to the mean squared transition dipole moment within the interval. In addition ε_0 is the vacuum permittivity, and σ^2 accounts for the variance introduced by neglected semiclassical vibrational effects. A scaling factor of 0.94 was applied to the MD-PMM excitation energies, following the approach adopted in previous works [32, 40]. This factor was chosen to match the first experimental absorption peak of lumiflavin, compensating for the DFT functional error.

Chapter 3

Results and Discussion

In this chapter, the two most significant research works conducted during the doctoral program are presented. In both studies, relevant biocatalytic systems were investigated and modeled using hybrid computational methods.

The first section focuses on a system of notable industrial relevance: the horse liver alcohol dehydrogenase (HLADH). This enzyme catalyzes the reversible reduction of ketones to *(S)*-alcohols with remarkably high yield and stereoselectivity. In our study [15] we proposed a possible reaction mechanism in the presence of the biomimetic cofactor BNAH (1-Benzyl-1,4-dihydronicotinamide). This study could potentially pave the way for the design of biocatalytic pathways involving the use of a biomimetic cofactor in the production of *(S)*-alcohols.

The second section is related to the fatty acid photodecarboxylase from *Chlorella variabilis* (CvFAP), the most recently discovered among the three known natural photoenzymes. Here we present an accurate simulation of the FAD absorption spectrum in CvFAP in the dark state and possible explanations for its red-shift experimentally observed later along the photocycle.

3.1 Modeling the HLADH-Catalyzed Formation of *(S)*-Alcohols

This section is devoted to the computational mechanistic investigation of the stereoselective reduction of prochiral ketones to chiral S-alcohols catalyzed by HLADH in the presence of a biomimetic cofactor.

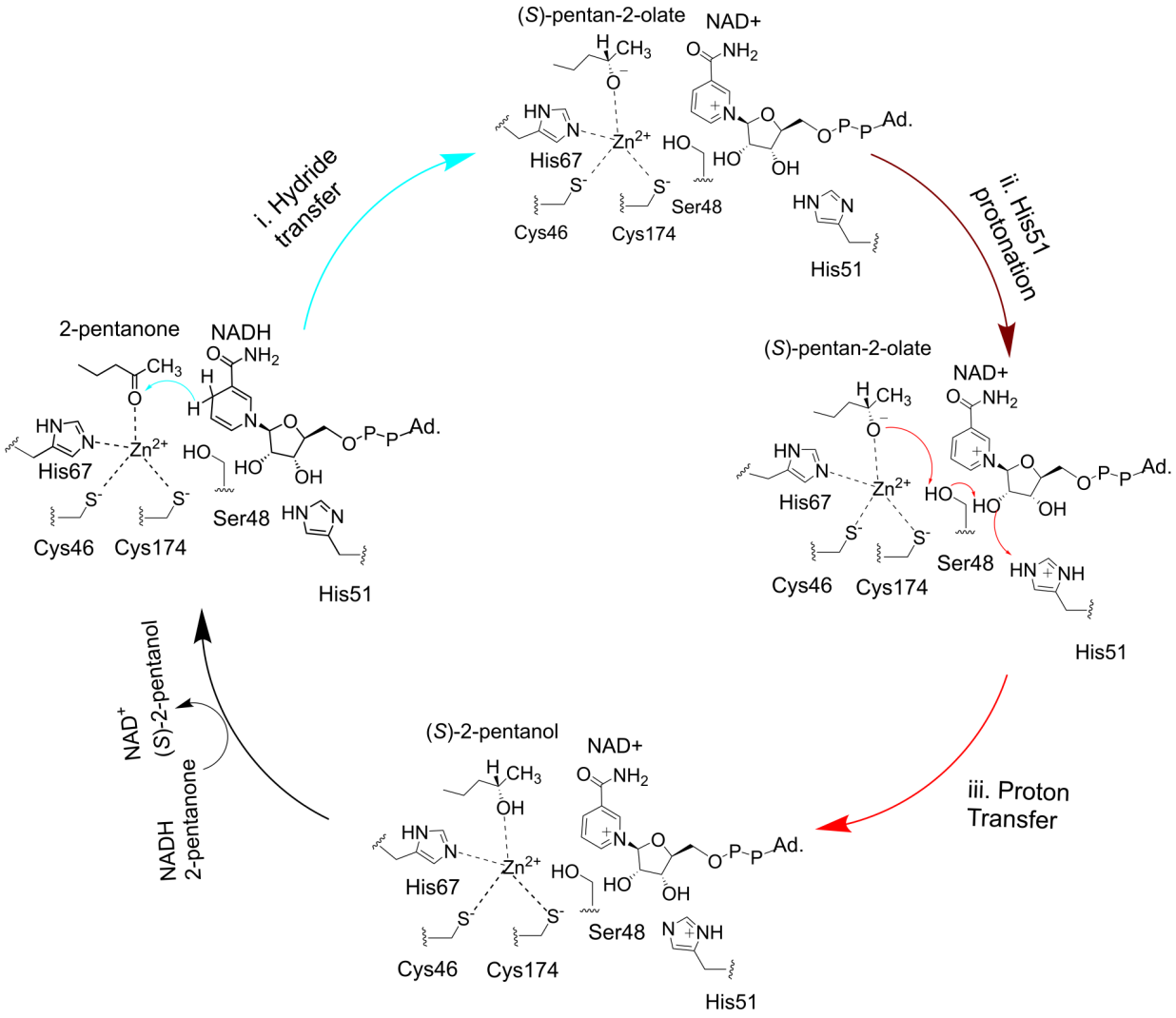


Figure 3.1: Reaction mechanism of 2-pentanone reduction catalyzed by HLADH in the presence of the native 1,4-NADH cofactor [111, 17, 15]. In the first step (i), a direct hydride transfer occurs from the C_4 atom of NADH to the C_2 atom of the substrate, forming the corresponding *(S)*-pentan-2-olate intermediate. In the following step (ii), His51 is protonated by the bulk. Subsequently (iii), a proton is transferred from $His51^+$ to the *(S)*-pentan-2-olate, yielding the corresponding *(S)*-alcohol. This step is mediated by a hydrogen-bond network connecting $His51^+$ to the *(S)*-pentan-2-olate through the riboside moiety of NADH.

According to the established mechanistic model (referred to as the canonical mechanism) of HLADH catalysis [112, 17], the formation of the reactive complex (Figure 3.1), in the presence of the natural 1,4-NADH, involves the coordination of the ketone carbonyl group to the Zn^{2+} ion. Simultaneously, the 1,4-NADH cofactor is positioned within the active site through multiple electrostatic and Van der Waals interactions, which orient the C_4–H bond close to the carbonyl group of the substrate. This geometry favours the hydride transfer (HT) step, in which the ketone is reduced to the corresponding (*S*)-alkoxide, while 1,4-NADH is oxidized to 1,4-NAD^+. In the following step His51 is protonated by bulk water and a proton relay system, involving His51, the ribose ring of 1,4-NADH, and Ser48, enabling the protonation of the alkoxide and the formation and release of the chiral (*S*)-alcohol.

Fish et al. [14] proposed a reaction scheme for the HLADH-catalyzed synthesis of *(S)*-alcohols from prochiral ketones, in which the natural cofactor NADH is replaced with biomimetic analogues. Among these, 1-benzyl-1,4-dihydronicotinamide (BNAH) has emerged as the most effective. For simplicity, as reported in figure 3.2 , we will denote the biomimetic cofactor BNAH as **1** and its oxidized form (BNA^+) as $\mathbf{2}^+$ as reported in figure 3.2.

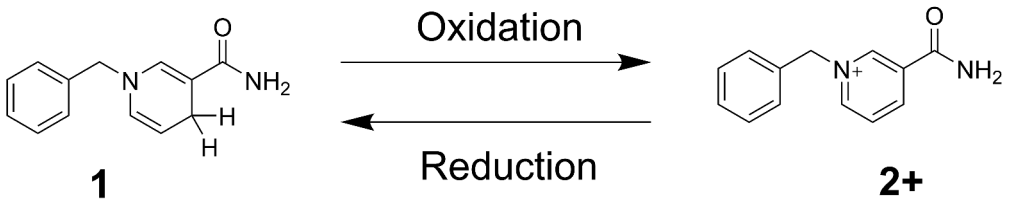


Figure 3.2: Molecular structures of BNAH (1-benzyl-1,4-dihydronicotinamide) in its reduced (1) and oxidized (2^+) forms.

According to the reaction cycle proposed by Fish et al., the biomimetic could be introduced in its oxidized form (BNA^+) and subsequently reduced in situ to BNAH using the rhodium complex $[Cp^*Rh(bpy)(H)]^+$. The resulting BNAH would then participate in the catalytic cycle of HLADH, reducing the ketone to the corresponding *(S)*-alcohol, and it is regenerated in its oxidized form at the end of the cycle, ready for reuse. This proposed "tandem catalysis" process could significantly lower process

costs while offering an efficient, straightforward, and environmentally friendly solution.

The experimental findings suggest that the smaller size of the biomimetic cofactor may lead to altered interactions with the enzyme-substrate complex, and alternative hydride transfer pathways deviating from the one proposed for the natural cofactor could exist. Furthermore, the proton relay mechanism responsible for the proton transfer (PT) step following hydride transfer, illustrated in Figure 3.1, is not viable in the absence of the ribose ring, which is not present in the biomimetic structure [15].

Collectively, these results support the existence of an alternative mechanistic pathway that enables the enzyme to maintain high catalytic efficiency even when 1,4-NADH is substituted with a structurally simplified biomimetic cofactor.

The first alternative mechanistic hypothesis was formulated by Marrone and Fish [16] by means of *ab initio* calculations (DFT) on a reduced active site model in gas-phase and it is reported in figure 3.3. The authors proposed a mechanism that starts with the His67 detachment from Zn^{2+} ion, subsequently, the system evolves toward the formation of an η^2-5,6-N-benzyl-1,4-dihydronicotinamide-Zn(II) complex (η^2-complex). In this configuration, hydride transfer is sterically favored and is rapidly followed by a proton transfer step. However, the mechanism of the proton transfer was not fully clarified. Also, the study did not explicitly account for environmental effects, which often play a relevant role in protein systems.

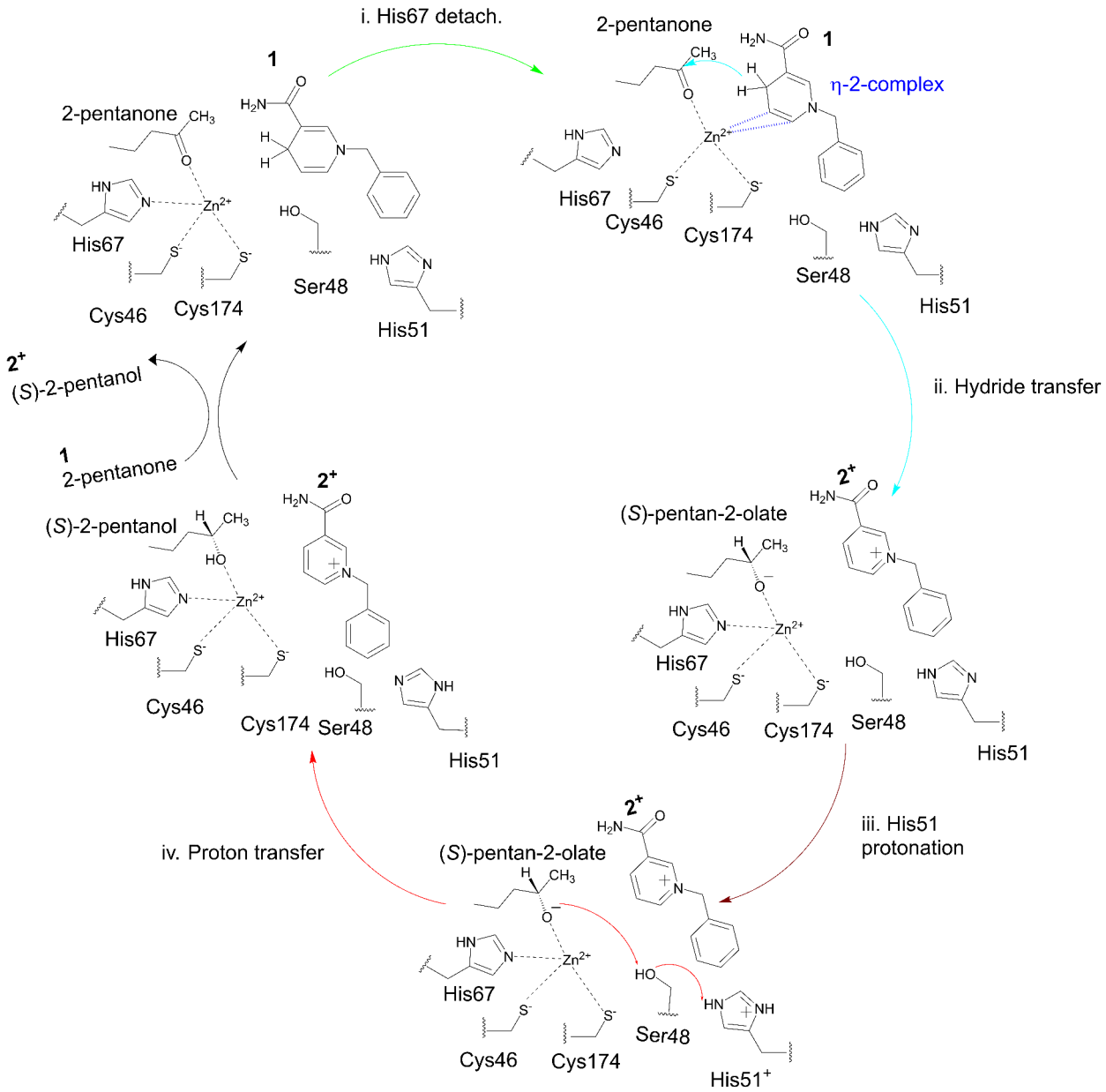


Figure 3.3: Alternative mechanism proposed by A. Marrone et al. in the presence of BNAH biomimetic [15, 16]. The proposed mechanism involves an initial dissociation of His67 from the Zn(II) center (i) yielding the η^2-5,6-1,4-dihydronicotinamide–Zn(II) complex (η^2-complex). The formation of this intermediate is followed by a direct hydride transfer from C_4 of BNAH to C_2 of the substrate (ii). The catalytic cycle proceeds with the protonation of His51 by the bulk solvent (iii), which allows the final proton transfer step (iv) required to form the (*S*)-alcohol.

Taking into account the previous computational results reported by Marrone et al. [16] and the experimental work of Fish et al. [14], here the HLADH-catalyzed reduction of 2-pentanone to the 2-*(S)*-pentanol in the presence of the biomimetic cofactor BNAH, was investigated using an hybrid computational approach which relies on both

MD-PMM and QM/MM calculations. The aim was to identify the most plausible reaction mechanism and to test the feasibility of the mechanistic hypotheses proposed by Marrone and Fish [19, 16].

The main difference between the canonical and alternative mechanisms is the hydride transfer step. In the canonical pathway, the hydride is directly transferred from the C_4–H position of 1,4-NADH to the ketone substrate. In contrast, the alternative pathway involves two steps for the hydride transfer: (i) His67 decomplexation, leading to the η^2-complex involving BNAH and the catalytic Zn^{2+} ion, and (ii) the proper hydride transfer, resulting in the formation of the *(S)*-pentan-2-olate and the BNA^+. In both mechanisms, the hydride transfer is subsequently followed by a proton transfer step, which completes the formation of *(S)*-2-pentanol.

Considering the experimental and computational evidences reported by several authors [16, 17, 111, 113, 114], it becomes clear that His51 plays a crucial role. Its charge state, governed by the protonation of the imidazole side chain, was therefore considered either as neutral (protonated at Nϵ, referred to as HIE51) or as cationic (doubly protonated at both Nϵ and Nδ, referred to as HIP51). The two forms represent the inactive and the active forms of the proton relay system, respectively.

An open question concerns the proton transfer mechanism in the presence of the biomimetic cofactor. In natural 1,4-NADH, the proton transfer chain is mediated by its ribosyl ring, which, however, is absent in the biomimetic BNAH. Nevertheless, HLADH can still retain its catalytic activity also with BNAH. One possibility is that the proton transfer phase proceeds via an alternative mechanism, either through a reorganization of the active site enabling direct transfer between His51 and Ser48, or through the involvement of another residue compensating for the absence of the ribosyl ring. A possibility is that one or more solvent molecules (water) activate the proton transfer chain by establishing a hydrogen-bond network connecting His51 and Ser48.

3.1.1 Molecular Dynamics simulations

For this investigation we adopted the computational workflow already reported in figure 2.3. To sample the conformational space of the system a MD simulation of HLADH with the substrate (2-pentanone) and the biomimetic cofactor BNAH was conducted for 50 ns, considering a neutral His51 state, with the substrate inserted in the active site with a pro-*S* orientation and with the BNAH biomimetic replacing the natural 1,4-NADH of the crystallographic structure . The choice of the specific chirality was simply based on the observation [14] of an enantiomeric excess of around 85% of the *(S)*-form for this specific reduction reaction in the presence of the same biomimetic cofactor. A representative structure of the HLADH active site is shown in Figure 3.4 (a). The MD simulation showed a stable active site, with a low RMSD of the protein backbone and no significant changes in the distance between the C_4 atom of BNAH and the C_2 atom of 2-pentanone, which consistently remained at approximately 4.2 Å as reported in figure 3.4 (b-c). Moreover, a visual inspection of the simulation confirmed that the substrate also retained the prochiral *(S)*-configuration and the Zn^{2+} ion maintained a stable tetrahedral coordination. The C_4(BNAH)-H_4-C_2(2-pentanone) angle (figure 3.4 (d)) remains near the ideal value of 180^o during the whole simulation with only little variations.

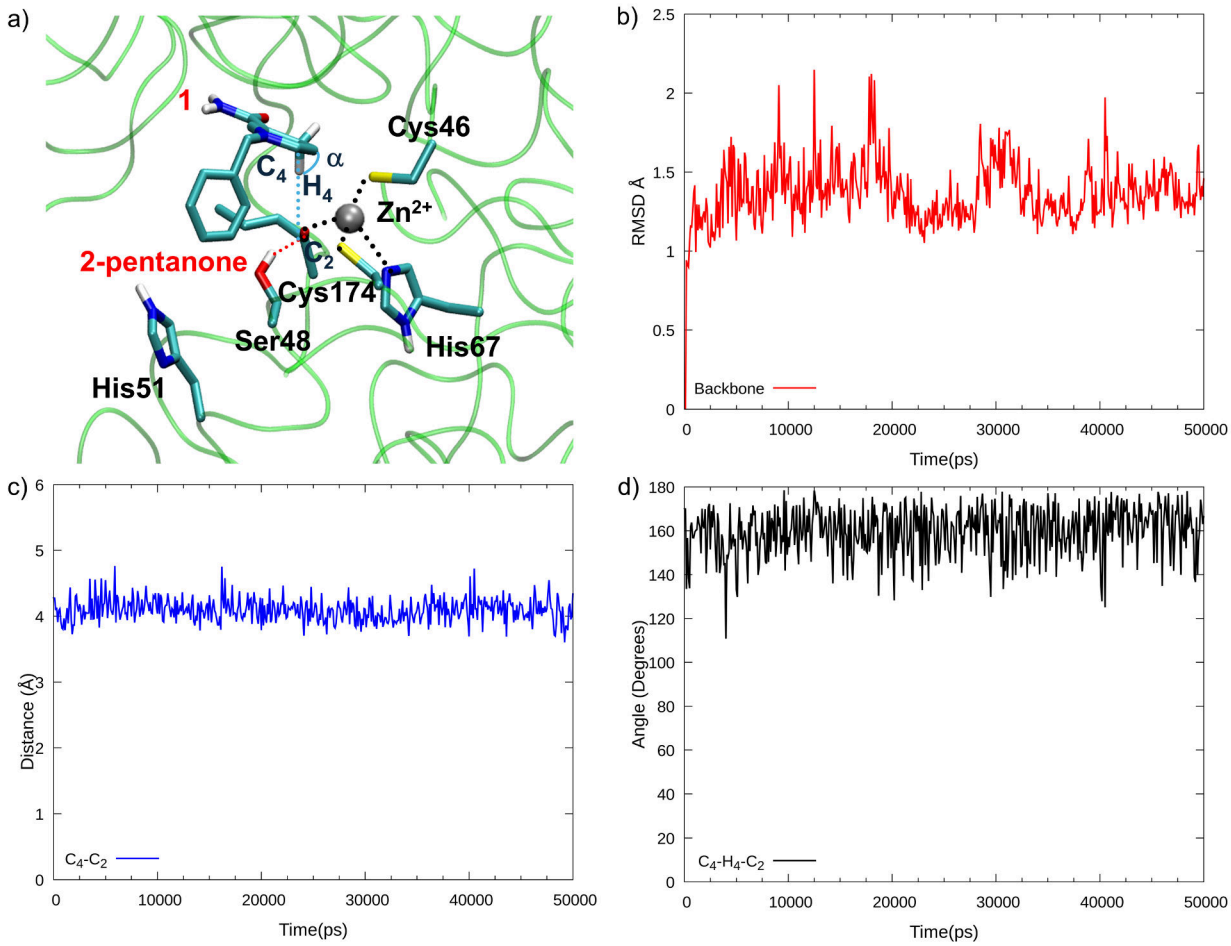


Figure 3.4: a) Active site of HLADH in the HIE51 configuration. The catalytic Zn^{2+} ion is coordinated by Cys174, Cys46, His67, and the oxygen atom of 2-pentanone. The 2-pentanone substrate is positioned in a pro-*S* orientation, while the natural 1,4-NADH cofactor is replaced by BNAH (denoted as 1). Here the hydrogen bond between Ser48 and the carbonyl oxygen of 2-pentanone is shown with red dotted line, the C_4-C_2 distance is highlighted in blue and the $C_4 - H_4 - C_2$ angle is denoted as α. The coordination of the catalytic Zn^{2+} is depicted with thick black lines. b) RMSD (Root Mean Square Deviation) calculated on the protein backbone. c) Time evolution of the C_4(BNAH)-C_2(2-pentanone) distance. d) Time evolution of the C_4(BNAH)-H_4-C_2(2-pentanone) angle.

In our study we will examine two different reaction mechanisms: the *alternative* [16] and the *canonical* [17, 111, 112]. In the alternative mechanism the process is strictly sequential: His67 decomplexation is followed by spontaneous formation of the η^2-complex, which, in theory, should enable an almost barrierless hydride transfer and a subsequent proton transfer. For the canonical mechanism we propose two distinct pathways: the sequential (or asynchronous) pathway, in which HT and PT occur in two separate steps, and the concerted (or synchronous) pathway, where HT and PT

occur simultaneously in one step.

3.1.2 Mechanistic investigation of the HT step

In this section we present an investigation of different reaction mechanisms through computed geometric pathways and corresponding potential energy surface (PES) scans with a QM/MM (ONIOM) approach [42, 77].

The PES scan is a well-known computational technique used to investigate the energetic profile of a molecular system as a function of one or more reaction coordinates. By gradually varying a small set of constrained coordinates, while optimizing the remaining degrees of freedom, the resulting energy profile provides a discrete representation of the potential energy surface corresponding to the chemical process associated with changes in the selected coordinates.

This approach, particularly valuable for comparing different possible reaction pathways, allows for the identification of geometries representing the highest energy point (usually very near to the transition states). In this way, it provides both qualitative and quantitative insights into the reaction mechanisms, including estimates of the energy barriers associated with the different pathways. When conducted in a QM/MM framework, PES scans also take into account the influence of the protein environment on the energy barrier.

As consistently highlighted in previous studies, Zn^{2+} is tetracoordinated [17, 19, 111], hence all zinc-coordinating residues together with the cofactor and the substrate, must be included within the QM region. Furthermore, experimental evidence has demonstrated that Ser48 and His51 play an active role in the proton transfer step [111, 114]. Consequently, the entire active site was fully incorporated within the QM region.

Due to the large size of the active site and the large number of geometry optimizations required for each PES scan, the QM region was treated at the GFN2-xTB level [57] and includes: Cys46, Ser48, His51, His67, and Cys174 side chains, the substrate

(2-pentanone), the biomimetic cofactor, the catalytic Zn^{2+} ion, and in some cases the water molecule bridging His51 and Ser48. The remaining part of the system was modeled using the GFN-FF force field [72]. The GFN-FF force field was chosen because, according to Grimme et al. [60], it is well-tuned for the description of very large systems such as metallo-proteins and metal-organic compounds. HLADH, that is an enzyme incorporating the Zn^{2+} in its active center, represents the exemplary application of this method.

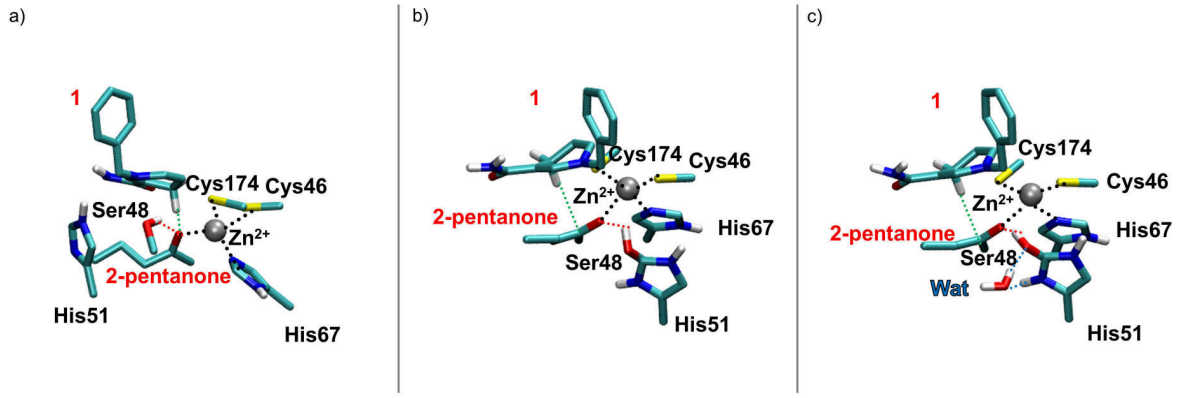


Figure 3.5: QM/MM optimized structures corresponding to the initial points used in the ONIOM scans are shown. The QM region in the ONIOM scans, corresponding to the entire HLADH active site, is depicted. Here all amino acids are saturated at the Cα atom (Ser48,His51,Cys174,Cys46) while the substrate (2-pentanone) and the cofactor (BNAH) are fully included in the QM Part. Three different configurations were tested: (a) Neutral His51 (HIE51); (b) Cationic His51 (HIP51); (c) Cationic His51 (HIP51) with an additional water molecule inserted between His51 and Ser48. Hydrogen bonds between Ser48 and 2-pentanone are highlighted in red, while those between water and other residues are highlighted in blue. The H_4(BNAH)-C_2(2-pent.) distance (reaction coordinate) for each structure is indicated with green lines: (a) 2.63 Å, (b) 2.59 Å, and (c) 2.52 Å. Structure reported in panel (a) correspond to the QC region in the subsequent MD-PMM calculation.

The starting point that we used for the subsequent PES calculations was extracted from the HIE51 MD trajectory. The choice of the specific frame was based on the key geometrical coordinates governing hydride transfer that are the distance between C_4−H of BNAH and the carbonyl carbon (C=O) of the 2-pentanone (d=4Å) and the C_4–H_4–C=O angle, which is close to 180°. These values provide a geometry that we assumed to be particularly favorable for the hydride-transfer step [19]. From this structure (see fig. 3.5 (a)) , other two distinct initial configurations, reported in fig 3.5 (b) and (c), were manually built and subsequently optimized using the ONIOM

algorithm (QM/MM) implementation present in the xTB software. The obtained geometries represent the starting point for six different PES scans, summarized in table 3.1.

Table 3.1: HLADH simulated PES scans. Models I and V starts from structure shown in Figure 3.5 (a); starting structure for model II is shown in panel (b) and the starting point for models III, IV, and VI is shown in panel (c).

Model[a]	Mechanism[b]	Pathway[b]	His51 charge
I	Canonical	Asynchronous	0
II	Canonical	Asynchronous	+1
III	Canonical	Asynchronous (+Wat1)	+1
IV	Canonical	Synchronous (+Wat1)	+1
V	Alternative	Asynchronous	0
VI	Alternative	Asynchronous (+Wat1)	+1

[a] Index of the simulated scan.
[b] Type of mechanism: Canonical denotes direct HT and Alternative is the mechanism proposed by Marrone et al. [16], which entails a detachment of His67 from the reaction center before hydride transfer. For canonical, two pathways are possible: synchronous, where HT and PT are concerted, and asynchronous, when PT follows HT. In some configurations (III, IV, VI), the bridge water molecule between His51 and Ser48 (denoted as Wat1) is present.

Models I and V start from the HIE51 configuration (structure **a** in Figure 3.5). In contrast, models II and III has a charged His51 residue, containing, respectively, no water molecule (Model II) or a single bridging water molecule (Model III), corresponding to structures **b** and **c** in Figure 3.5. For each starting configuration, the PES scan was performed using the distance between the C_2 atom of 2-pentanone and the H_4 of BNAH as the reaction coordinate. This distance was systematically reduced from its QM/MM-optimized value to 1.10 Å, which corresponds to the optimal C_2–H bond length in the *(S)*-pentan-2-olate coordinated to Zn(II).

With regard to the concerted pathways and, more broadly, to the configurations initiating with cationic His51 (HIP51), it should be emphasized that these configurations are already regarded as less probable, considering the previous studies. In the presence of the ketone, previous work that considered the enzyme with natural 1,4-NADH [115] showed that the pK_a of the N_δ of His51 within the active site is approximately 6.7 before the hydride transfer step. At pH 7, the same of R. Fish [14] experimental work, it is therefore very unlikely that this atom is in its HIP form prior to hydride transfer

at this pH.

Moreover, also the work of Plapp et al. [114] further supports the view that His51 is neutral before the hydride transfer step and exists in its cationic form only as a transient species during the proton-transfer step. Nevertheless, due to the absence of experimental data for the specific reaction in the presence of the biomimetic compound, these potential reaction pathways have been investigated.

Models V and VI represent the alternative mechanism, considering the HIE51 and HIP51 protonation states, respectively. This mechanism was investigated through two sequential scans. In the first relaxed scan, the length of the Zn–$N\delta$ (His67) coordinative bond was varied from 2.38 to 4.00 Å to simulate the detachment of His67 from the catalytic Zn^{2+} ion, leading to the η^2-5,6 C=C bond coordination of the biomimetic cofactor BNAH to the Zn(II) metal center. This scan was followed by the proper hydride transfer step, which was investigated using the same reaction coordinate of in the canonical mechanism.

For each scan, the "unperturbed energy barriers" were calculated. These were obtained by extracting, for each point of the PES scan, the QM/MM-optimized geometry of the reaction center (the QM region) at the GFN2-xTB/GFN-FF level of theory and then performing gas-phase DFT single-point calculations at the B3LYP/6-31G* level of theory on the isolated QC structure. This choice yields reliable coordination geometries at a modest cost within a QM/MM framework, while avoiding the known inaccuracy of semi-empirical methods for estimating transition state energies. Finally, the environmental contribution, which is omitted in the unperturbed single-point calculation, is reintroduced explicitly and dynamically through MD-PMM calculations. This procedure allows obtaining reliable geometries for the Zn(II) center, while also providing consistent electronic energy differences and statistically converged free-energy barriers for the hydride-transfer mechanism

The use of these unperturbed energies was also preferred over ONIOM energies, since in the ONIOM algorithm implemented in xTB electrostatic embedding is not present. As

already reported in recent literature [43], this omission has little impact on geometries but the energy differences between distinct geometries (e.g., reactant and highest-energy point) are highly sensitive to the polarization of the QM region and hence, without the electrostatic embedding, the xTB/GFN-FF ONIOM energy barriers are not accurate.

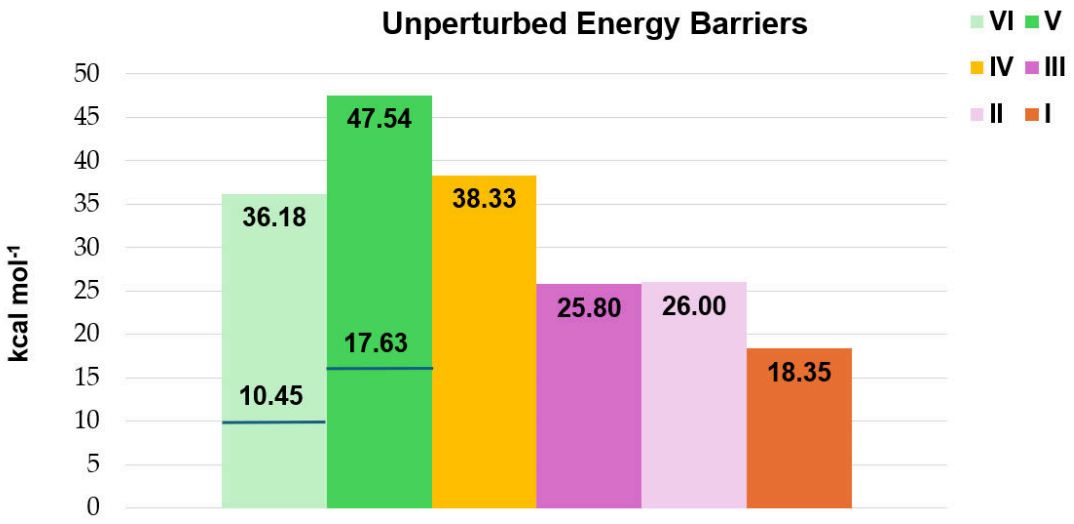


Figure 3.6: Unperturbed energy barrier along each mechanistic path calculated at B3LYP/631G* level of theory. The blue lines in the mechanisms V and VI indicate the value of the energy barrier for the histidine detachment step.

The unperturbed energy barriers, reported in figure 3.6 indicate a strong preference for the model I, while all configurations with cationic His51 are disfavored. All evidence thus suggests that the His51 is likely to be neutral in the hydride transfer state, in perfect agreement with previous work. These results also demonstrate that the alternative mechanism seems not viable: the total energy barrier computed for reaction models V-VI always exceeds 35 kcal mol^{-1} independently of the protonation state of His51 or of the presence of a bridging water molecule.

Building upon these reference data, Gibbs Free Energy barriers were calculated by means of MD-PMM, exclusively for reaction model I and for reaction model V, which originates both from the configuration with neutral His51.

3.1.3 Investigation of free energy barriers for HT step

A crucial element in all computational studies aimed at elucidating reaction mechanisms is the calculation of Gibbs' free energy barrier that represents the free energy barrier the reactants must overcome to convert into products. This quantity, denoted as ΔG^{o*} (Standard Conditions), is defined as the difference between the Gibbs free energy of the transition state, G_{TS}, and that of the reactants, G_{R}. In our case we assume that the transition state is very near the local highest energy point. We define, instead, as ΔG^{o} the Gibbs free energy difference between products and reagent structure.

In accordance with PMM principles, already reported in methods section, the unperturbed energies and dipoles of the Quantum Center (QC) ($\widehat{H_0}$), corresponding to the optimized QM region extracted from the ONIOM (reported in fig 3.5 (a)) , were calculated at the DFT level in the gas-phase for each of the crucial geometries at B3LYP/6-31G*. These are: the reactant (first point), the transition state-like structure (the highest energy point), and the product state (last point) of the hydride transfer step for the mechanism I. For mechanism V four geometries were considered: the reactant structure, the η^2-complex formed after the dissociation of the His67 from Zn(II) center, the transition-state-like structure associated with the hydride transfer (which corresponds to the highest energy point), and the product state.

To determine the most suitable PMM approach, particular attention was paid to the size of the QC region (approximately 90 atoms) and the extensive number of frames of the molecular dynamics simulation (45000). The relatively large QC region poses a considerable computational challenge. To address this, the dipole-based variant of the PMM method was adopted [32], restricted to the electronic ground state. The expression used to compute the perturbed Hamiltonian matrix $\widehat{H}$ is reported in Equation 2.6, while the corresponding calculation of the Gibbs free energy is calculated

from the diagonalization of $\widehat{H}$ at each MD frame according to the expression reported in Equation 2.7 in the Methods section.

Starting from the complete molecular dynamics trajectory, the initial 5000 frames (corresponding to 5 ns) were excluded to ensure proper system stabilization. The remaining part was divided into three intervals, each consisting of 15,000 frames (15 ns). The average Gibbs Free Energy for each structure, shown in Figure 3.7, was calculated as the standard mean across these three segments.

Our data show that the canonical mechanism sequential pathway (Model I) shows a late transition state, with a free energy barrier ΔG^{o*}= 14 kcal mol^{-1} in perfect agreement with the value previously reported by other authors [17, 112, 111] in the presence of the natural 1,4-NADH. The resulting product also seems strongly favored, showing a ΔG^{o} of just 1 kcal mol^{-1}.

The reaction profile for the alternative mechanism (Model V) shows instead a strongly endothermic reaction, with a first barrier of around 16 kcal mol^{-1}, for the His67 detachment and η^2-complex formation step, while the hydride transfer from the η^2 complex requires additional 28 kcal mol^{-1} (ΔG^{o*} = 44 kcal mol^{-1}). Finally also the product is disfavored showing a total ΔG^{o} of around 31 kcal mol^{-1}. The standard mean error on Gibbs free energy calculated by MD-PMM considering three different MD subsets is 1.80 kcal mol^{-1} for the model I and 2.40 kcal mol^{-1} for model V.

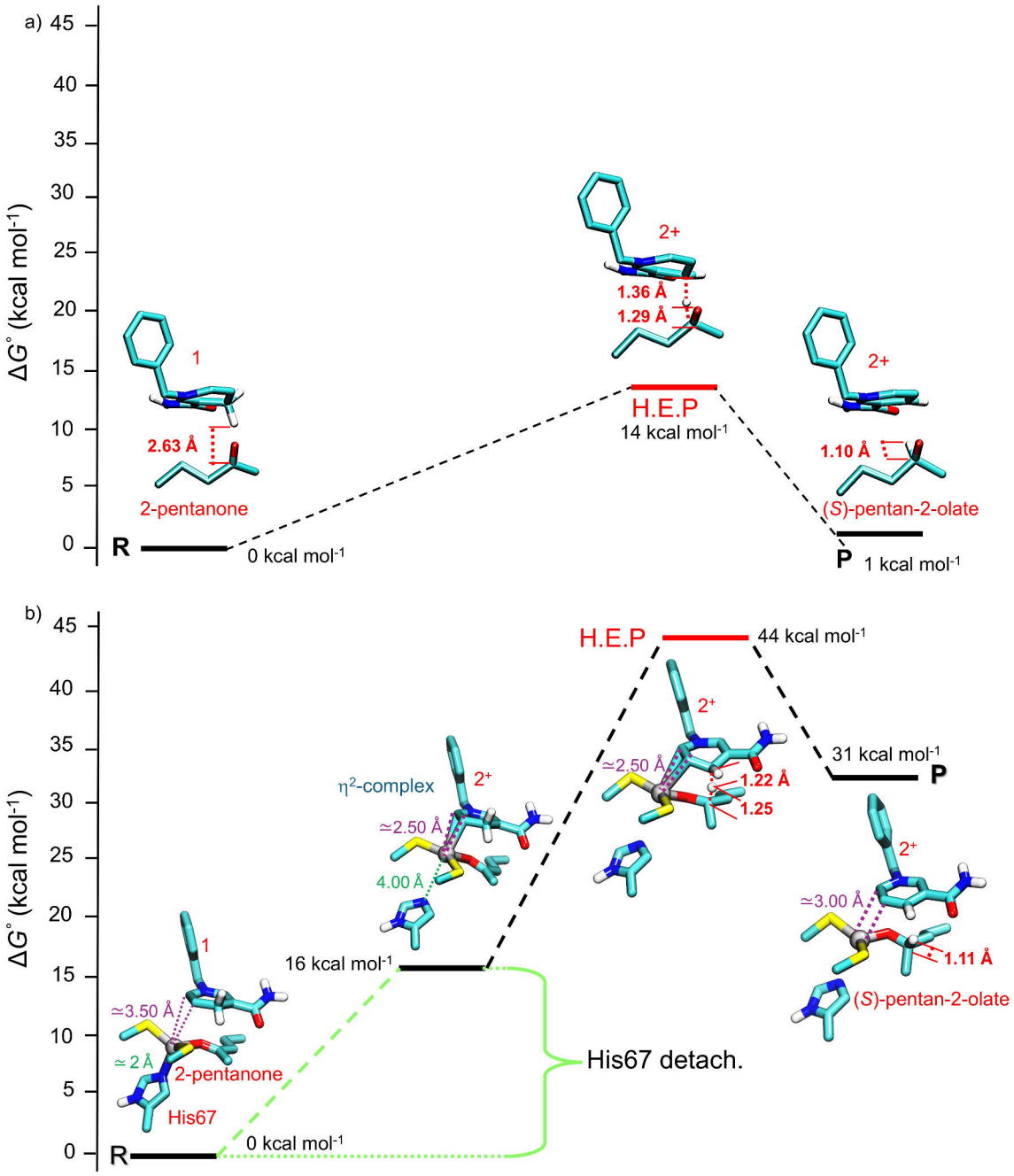


Figure 3.7: MD-PMM reaction pathways, comparing the Gibbs' free energy barriers of the reaction model I (a) and reaction model V (b). R represents the reagent structure, H.E.P. is the highest energy point, and P is the product structure. In the mechanism (b) also the geometry corresponding to the η^2-complex formed after His67 detachment is represented. Crucial bond distances, related to the reaction coordinates, for both reaction pathways are highlighted in red. In the mechanism (b) the His67 contribution to the energy barrier is highlighted by a green curly bracket.

The MD-PMM data and ONIOM scans rule out the alternative mechanism and support a sequential pathway driven by direct hydride transfer. As this step matches the one with natural 1,4-NADH, the absence of the ribosidic ring implies that only the proton-

transfer step must follow a different mechanism.

3.1.4 A possible pathway for the PT step

The primary objective of this study was to determine, with high precision and accuracy, the feasibility of the alternative reaction mechanism. Both QM/MM data and MD-PMM simulations clearly demonstrate that, when the environmental contributions and the conformational fluctuations of the protein are explicitly considered, this mechanism is strongly disfavored both from thermodynamic and kinetic point of view.

Simulations of potential reaction pathways in which His51 is in the cationic state prior to the HT step (Models II, III, IV) reveal a significant energetic penalty relative to this charged state (HIP51). These results rule out the possibility of a concerted mechanism involving simultaneous hydride and proton transfer. Instead, our QM/MM calculations support a sequential pathway, consistent with the one previously proposed by Plapp et al. [114] for the native 1,4-NADH cofactor, in which His51 is neutral during the HT step, it is protonated from the bulk afterwards, and subsequently participates in the PT step. This sequential mechanism appears to be favored also in the presence of the BNAH cofactor.

Hence, in the final part of this work, we propose a mechanistic hypothesis for this proton transfer step in the presence of the BNAH cofactor.

For this purpose a second MD simulation was performed using the same setup, but changing the protonation state of His51 from the HIE to the HIP form. Also this trajectory, as shown in figure 3.8, displays a low RMSD and stable key geometrical parameters (C_4(BNAH)-H_4-C_2(2-pentanone) angle near 180^o and C_4-C_2 distance near 4 Å).

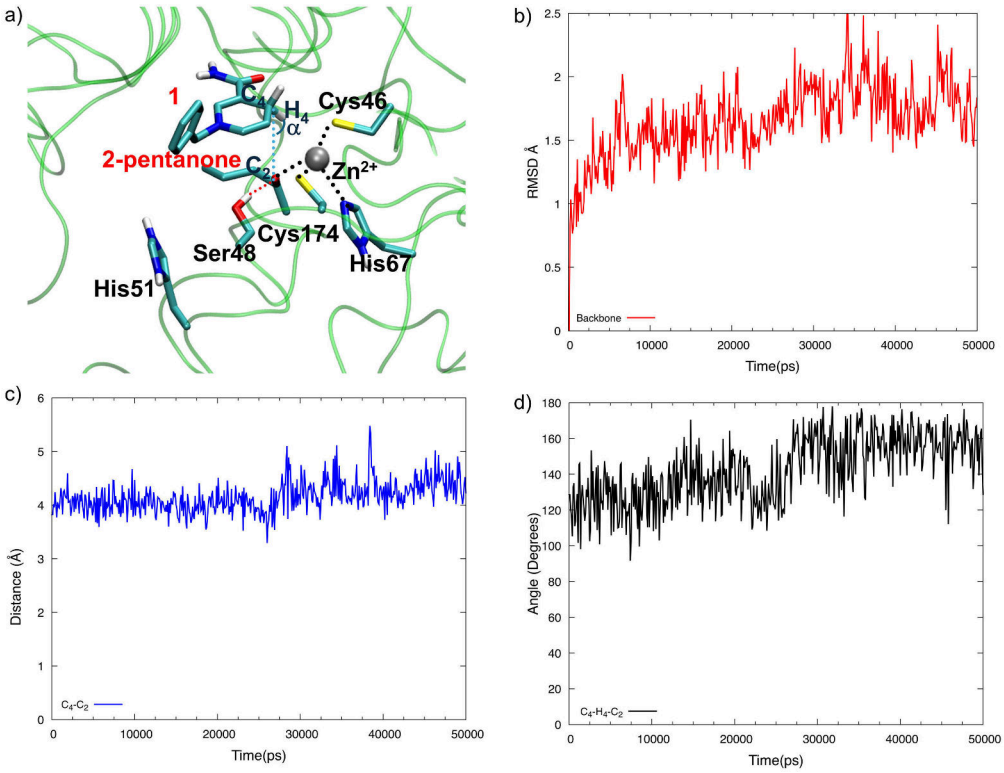


Figure 3.8: a) Active site of the HIP51 configuration. Here the hydrogen bond between Ser48 and the carbonyl oxygen of 2-pentanone is shown with red dotted line, the C_4-C_2 distance is highlighted in blue and the $C_4 - H_4 - C_2$ angle is denoted as α. The coordination of the catalytic Zn(II) is depicted with thick black lines. b) RMSD (Root Mean Square Deviation) calculated on the protein backbone. c) Time evolution of the C_4(BNAH)-C_2(2-pentanone) distance. d) Time evolution of the C_4(BNAH)-H_4-C_2(2-pentanone) angle.

To assess the feasibility of a water-mediated mechanism, the complete MD trajectory was examined using an in-house Fortran tool specifically developed for hydrogen-bond network analysis. This tool enabled the systematic identification of all trajectory frames in which one or more water molecules establish a hydrogen-bond network connecting His51 and Ser48. Our investigation demonstrated the presence of one stable water molecule linking N_ε(His51) and the OH group of Ser48 in approximately 2.5% of the frames (see figure 3.9 (a). Additionally, as reported in 3.9 (b), 8% of the structures exhibited multiple water molecules forming a hydrogen-bond network between the same atoms, enabling a water-mediated proton-transfer chain.

Building on these insights, and to verify our hypothesis, an ONIOM PES scan was conducted considering the simplest case (one bridging water molecule). Starting from the product structure obtained at the end of the PES scan corresponding to the hydride-

transfer step, in which the *(S)*-pentan-2-olate is formed, the geometry was subsequently optimized. From there the O–H distance between Ser48 and the oxygen atom of the *(S)*-pentan-2-olate was then selected as the reaction coordinate and varied from 1.50 Å (its value in the optimized structure of the *(S)*-pentan-2-olate) to 0.90 Å, corresponding to the typical O–H bond length observed in *(S)*-2-pentanone (see figure 3.9 (c)).

The results fully support our hypothesis: the reaction is barrierless and highly exothermic (ΔE_{R-P}=-15.67 kcal mol^{-1}), and the scan indicates an immediate, spontaneous water-mediated proton transfer from His51 to Ser48 that occurs varying only the reaction coordinate.

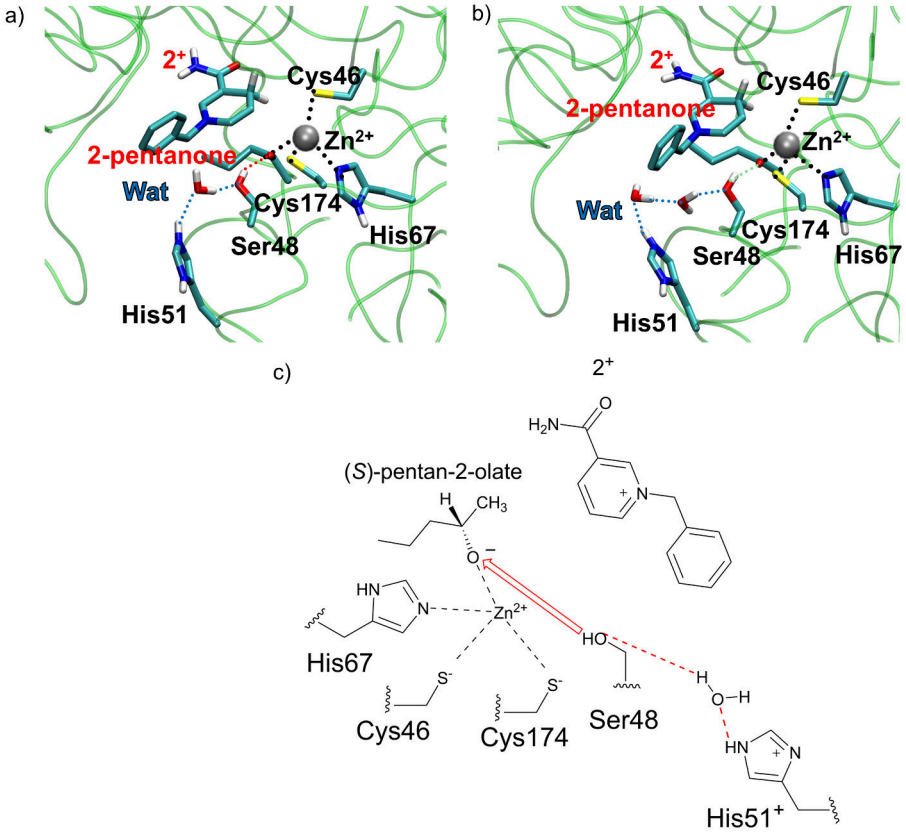


Figure 3.9: a) Reaction center of HLADH with charged His51 (HIP51). This snapshot from MD clearly shows a water molecule (Wat.) bridging His51 and Ser48, whose interactions are highlighted by blue lines. b) The hydrogen bond network involving two water molecules. c) Scheme of the ONIOM scan for the simulation of the proton-transfer chain. The reaction coordinate H(Ser48)–O(*(S)*-2-pentan-2-olate) is indicated by a red arrow, while the dotted red lines highlights the hydrogen bond network.

3.1.5 Our mechanistic interpretation and future perspectives

In conclusion, the canonical asynchronous mechanism appears to be the most favored, with a Gibbs free energy barrier of approximately 14 kcal mol^{-1} for the hydride transfer step, as calculated by MD-PMM, followed by a strongly exothermic proton transfer that complete the formation of the *(S)*-alcohol. The alternative mechanism, instead, seems not viable since the hydride transfer step involves a Gibbs free energy barrier of 44 kcal mol^{-1}.

According to our calculations, the most likely pathway is illustrated in Figure 3.10. The first step is the direct hydride transfer from BNAH to 2-pentanone, leading to the formation of the *(S)*-pentan-2-olate. In this step, His51 is initially neutral and is protonated only after the formation of the *(S)*-pentan-2-olate.

The following proton transfer step, which completes the formation of the *(S)*-alcohol, is activated by the presence of one or more water molecules that enable the proton transfer from His51 to Ser48 through a hydrogen-bond network. Finally, the Ser48 proton is transferred to the carbonyl oxygen concluding the formation of the alcohol.

In conclusion, this study is not only significant for the results obtained but also for demonstrating the capability of our computational approach, that integrates QM/MM and MD-PMM simulations in a sequential manner. The integration of the accuracy provided by QM/MM calculations with the statistical sampling of MD-PMM simulations proved crucial for achieving a comprehensive understanding of the reaction mechanisms in complex catalytic systems

.

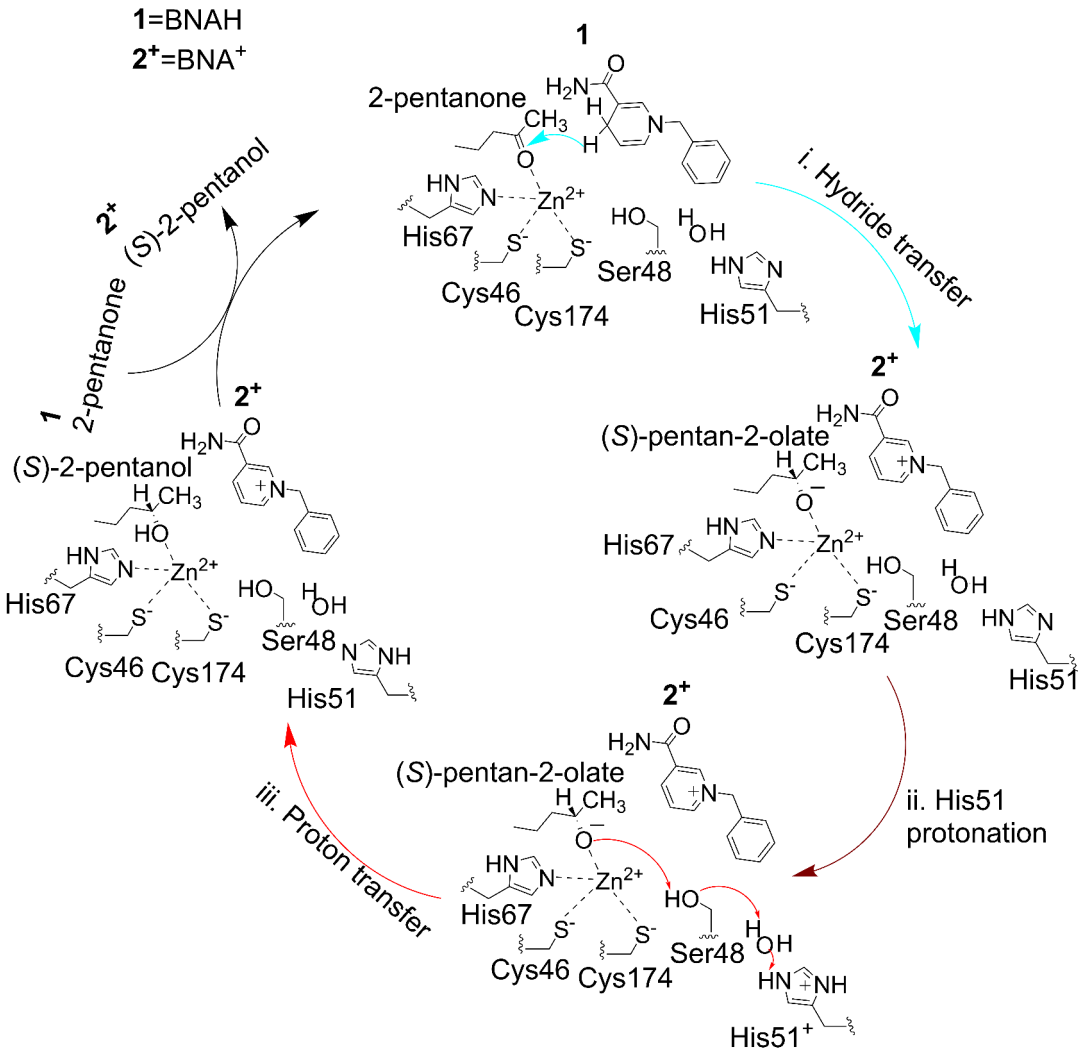


Figure 3.10: Proposed mechanism for the HLADH-catalyzed reduction of 2-pentanone with the biomimetic co-factor, BNAH. In the first step (i), a direct hydride transfer occurs from the C_4 atom of BNAH to the C_2 atom of the substrate, forming the corresponding *(S)*-pentan-2-olate intermediate. In the following step (ii), His51 is protonated by the bulk. Subsequently (iii), a proton is transferred from $His51^+$ to the *(S)*-pentan-2-olate, yielding the corresponding *(S)*-alcohol. This last step is mediated by a a water molecule connecting $His51^+$ to Ser48 through an hydrogen bond network.

3.2 Mechanistic and spectroscopic study of fatty acid photodecarboxylase

Fatty acid photodecarboxylase (FAP), DNA photolyase and protochlorophyllide reductase are the only three known photoenzymes. The former has attracted significant

interest due to its ability to convert long-chain fatty acids (e.g., palmitic acid) into hydrocarbons (RH) upon light excitation with a process involving a charge-transfer between the fatty acid and the oxidized flavin adenine dinucleotide (FAD) cofactor.

The full mechanistic pathway of this reaction is still not entirely understood. According to the mechanism proposed by Sorigué et al. [116], the reaction begins with a rapid light-driven step in which the electronically excited FAD accepts an electron from the deprotonated fatty acid ($RCOO^-$), forming two radical species: $FAD^{\bullet -}$ and $RCOO^{\bullet}$. This is followed by a rapid decarboxylation: $RCOO^{\bullet} \rightarrow R^{\bullet} + CO_2$. Subsequently, a back electron transfer from $FAD^{\bullet -}$ to $R^{\bullet}$ occurs, coupled with a proton transfer from an as-yet-unidentified donor (Cys432 and Arg451 were proposed [116, 97, 117]), completing the formation of the alkane. This step is linked with the formation of a more UV-Vis red-shifted species called FAD_{RS}. According to their hypothesis [116], the formation of FAD_{RS} could be attributed to two combined effects: alterations in the charge distribution within the active site and the formation of a hydrogen bond between FAD and a catalytic water molecule (Wat1).

Wu et al. [118] have considered the possibility of an alternative mechanism, where the substrate is in the neutral form (RCOOH) in the active site of the protein before the photoinduced reaction. The proposed mechanism, mediated by water molecules, does not require a residue that act as proton donor.

The absorption spectrum of CvFAP, in its resting state, was studied by several authors: Wu et al. [118] measured the absorption spectrum of the protein in the *apo* state (without the substrate), while Heyes et al. [117] reported the spectrum of the *holo* form (in the presence of the substrate). Only a minor difference is observed between the two spectra, as shown in Figure 3.11 and table 3.2. Both spectra show two main absorption peaks at 467 nm and 390 nm. The 467 nm peak is red shifted (10–15 nm) compared to typical flavoproteins [119, 120] that generally show this absorption in the range of 450-459 nm (highlighted in green). Compared to lumiflavin in water, instead, the position of the peak is shifted by 20 nm [121, 116].

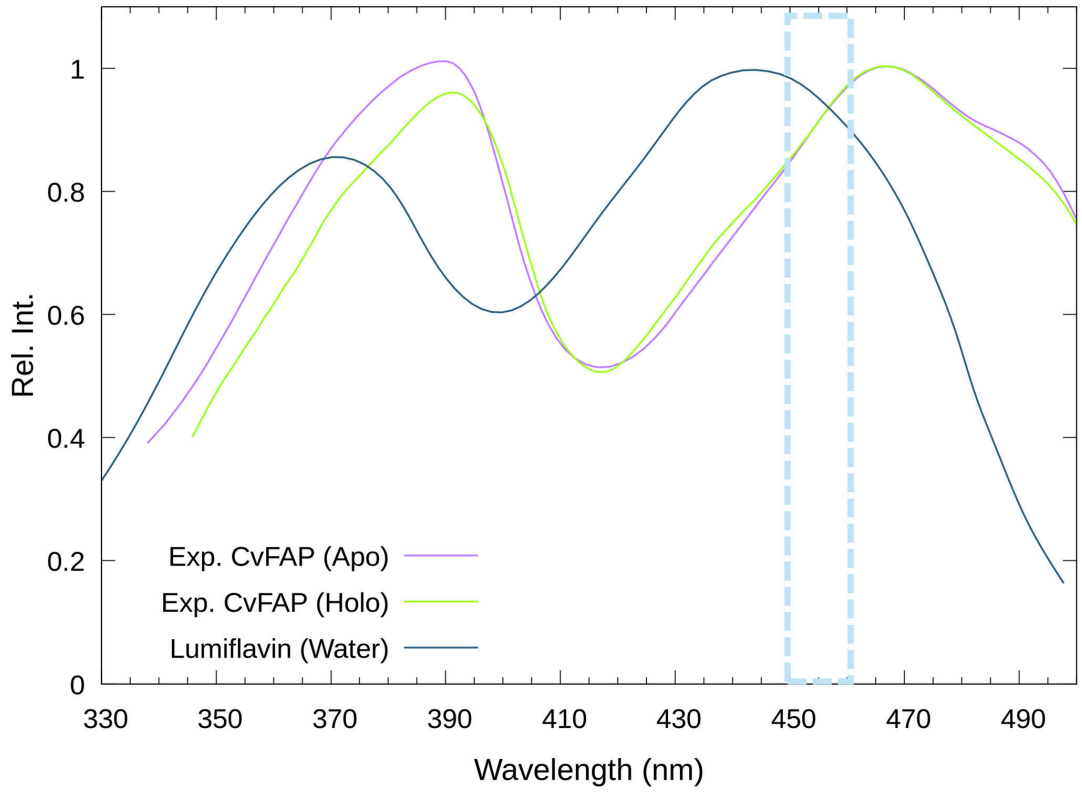


Figure 3.11: Experimental spectra of FAD in *apo* (purple) [118] and *holo* (green) [116] CvFAP compared to the experimental spectrum of lumiflavin in water (blue) [121]. The two FAD main peaks in CvFAP, at 467 and 390 nm, are red-shifted by 21-23 nm compared to the peaks of lumiflavin in water (443 and 370 nm). The cyan box highlight the spectral range (450–460 nm) where generally the first FAD absorption peak is observed in other flavoproteins.

Table 3.2: a) Positions of the main UV-Vis absorption peaks of FAD within *apo* CvFAP [118] and *holo* CvFAP [117] and of lumiflavin in water [121]; and (b) their relative intensity.

Reference	λ_1(nm)[a]	λ_2(nm)[a]	I_{rel}[b]
CvFAP *apo* [118]	467	389	1.01
CvFAP *holo* [117]	467	391	0.96
Lf. in water [121]	443	370	0.86

[a] Position of the principal UV–Vis absorption peaks, defined as the band center in the experimental spectra.
[b] Relative intensity of the two main peaks ($I_{\lambda_1}/I_{\lambda_2}$).

Previous studies conducted by Sorigue et al. [116, 97] revealed that the oxidized FAD cofactor (resting state) in CvFAP adopts a distinctive "butterfly-bent" conformation where the the C_4–N_5–N_{10}–C_9 dihedral angle (see figure 1.5 for atom numbers) is around $12-15°$. This bending has been previously observed in reduced flavin states, but it is uncommon in the oxidized form [118, 116]. Currently, the literature does not provide

a quantitative distinction between the contributions to the observed spectral red-shift arising from FAD bending and those related to the environmental effects.

In this thesis, we applied our computational workflow, combining QM/MM and MD-PMM calculations, to investigate the electronic properties of FAD in CvFAP. Unlike the previously analyzed system, the MD-PMM calculations here include excited states, while DFT methods are employed in the QM region.

3.2.1 Active-Site Protonation state in Fatty Acid Photodecarboxylase

The first part of this study is focused on the analysis of the excited states of Cv-FAP. Starting from the crystal structure (PDB code: 6YRU) [97], two models were constructed: one containing palmitic acid (RCOOH) and the other containing the corresponding palmitate anion ($RCOO^-$).

These structures were optimized in a QM/MM framework using ORCA 6. The QM region, treated at the B3LYP/def2-TZVP level, includes the head of the palmitic acid (-CH_2-CH_2-CO_2^- fragment), the lumiflavin moiety of FAD(N_5), the two catalytic water molecules (Wat1), and the side chain of Arg451. For the MM region the AMBER force field was used. The selection of residues in the QM region was motivated by previous studies that identified these atoms as possibly actively involved in the photoinduced reaction mechanism. [116, 117]. Optimized structures are shown in fig. 3.12. In both protonation states, CvFAP adopts a bent configuration with a bending angle of approximately 20°.

A key difference emerges between the two protonation states: in the neutral model (RCOOH), the catalytic water molecule Wat1 bridges FAD(N_5) and the backbone carbonyl oxygen of Ser574 through a hydrogen bond, conversely, in the deprotonated model ($RCOO^-$), the hydrogen bond is formed between the palmitate oxygen and the backbone carbonyl oxygen of Ser574.

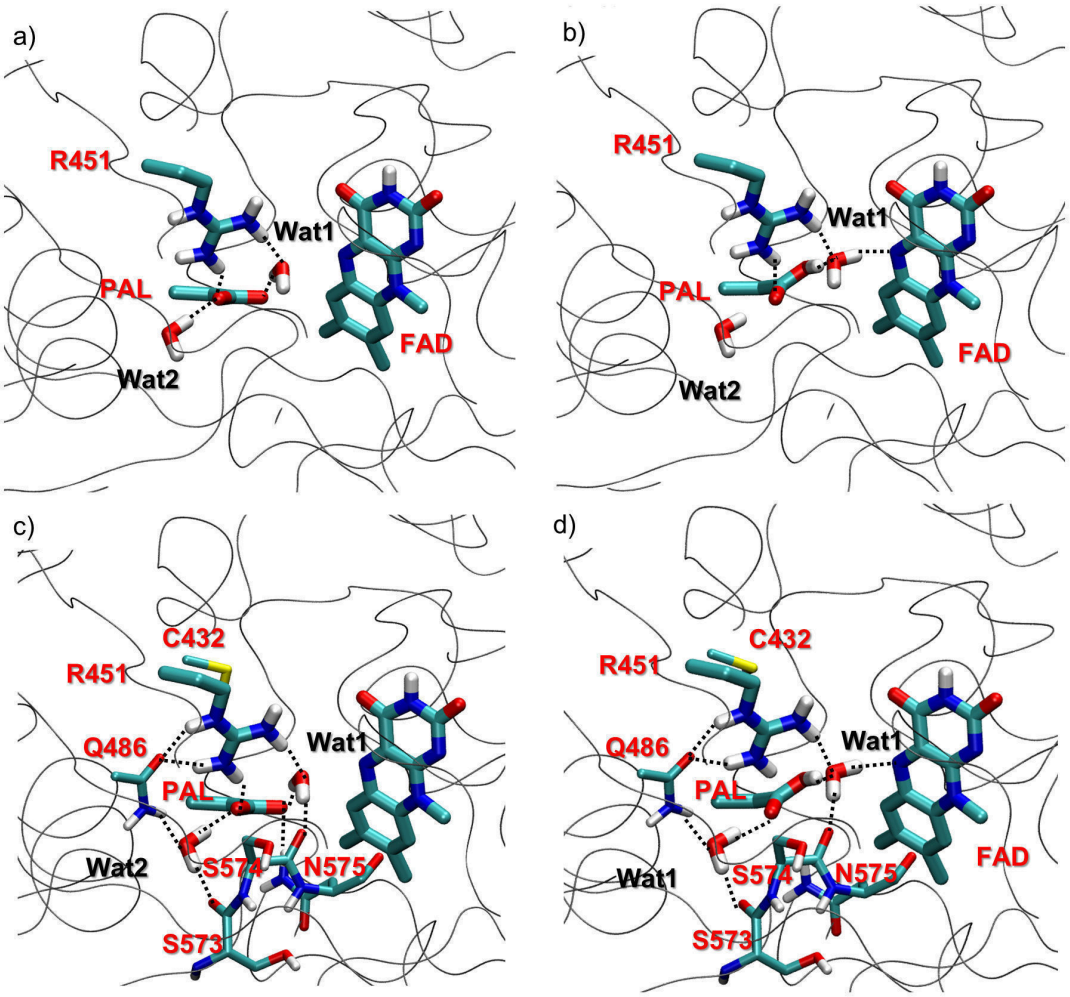


Figure 3.12: QM/MM-optimized structures of the CvFAP active site for the two substrate protonation states: QM region (a) and active site (c) of the deprotonated palmitic acid (palmitate) configuration; (b) QM region and active site (d) of the neutral palmitic acid configuration. The QM region includes: the head of the palmitic/palmitate moiety, the lumiflavin, two catalytic water molecules (Wat1 and Wat2) and the side chain of R451. The active site depiction highlights also the MM residues Q486 (side chain), C432 (side chain), S573, S574, and N575 that establish direct h-bond with QM atoms.

From the optimized geometries, the first 25 lowest-energy singlet electronic states were computed within the QM/MM framework at the TD-DFT level. For these calculations, we employed the range-separated CAM-B3LYP functional in conjunction with the def2-TZVP basis set; this choice reflects the known superiority of range-separated functionals in describing charge-transfer (CT) states [21].

The characterization of the charge-transfer states was carried out through transition density analysis using the Theodore Code [122]. The results were further corroborated by visual inspection of the natural transition orbitals (NTOs) [123]. Energies, oscillator

strengths, and character of the first eight excited states are reported in tables S1 and S2 in the appendix. In the palmitate configuration, the calculations locate the first charge transfer (CT) state at 4.15 eV (NTOs shown in Fig. 3.13 b), while the CT state was found at 6.31 eV in the palmitic acid configuration. We compared these results with those of Dell'Orletta et al. [124], who used the same QM/MM protocol for both geometry optimization and TD-DFT calculations but with the wB97X-D3/6-31+G(d,p) level of theory. Their findings are analogous, excluding possible artifacts arising from the choice of DFT functional.

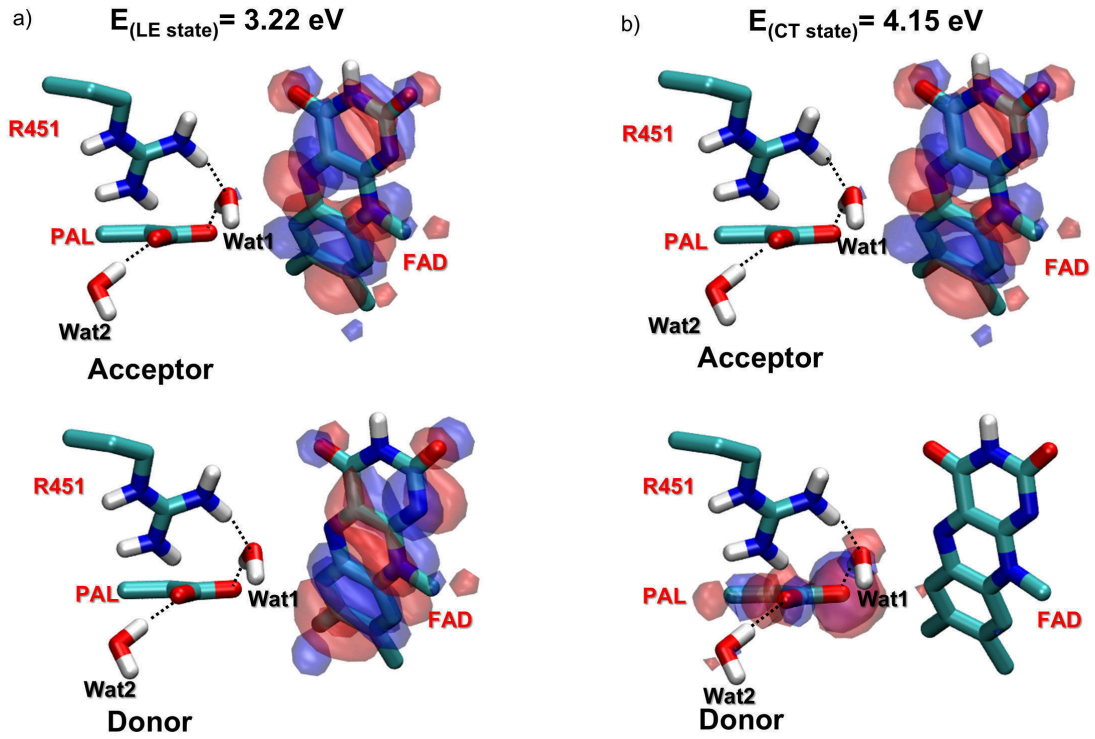


Figure 3.13: NTOs of respectively locally excited (LE) (a) and (b) (CT) state for the configuration containing the palmitate substrate. Calculations were performed at the CAM-B3LYP/def2-TZVP level of theory. Here only QM atoms, in the ground-state geometry, are depicted.

Finally, for the palmitate configuration, the LE and CT excited-state structures were optimized at the same level of theory and reported in figure 3.14. Comparing the optimized active-site geometries of LE and CT it is possible to observe a deep reorganization of the hydrogen-bonding interactions involving the palmitate ligand (PAL) and FAD. This change is favoured by PAL head rotation in the CT state. The most

significant differences regard the position of R451, which shifts slightly away from the PAL O_1 atom (N-O distance varies from 2.82 Å in the LE state to 3.70 Å in the CT state), thereby weakening the corresponding hydrogen bond. The Wat1 molecule moves away from the palmitate oxigen (O(Wat1)-O_2(PAL) distance is 2.90 Å in the LE state and 3.68 Å in the CT state), breaking its hydrogen bond with the palmitate and instead interacting with the N_5 atom of FAD. By contrast, the change in the FAD bending angle is, in our assessment, small and of negligible significance (from 20° in the ground state to approximately 24° in CT and 23° LE).

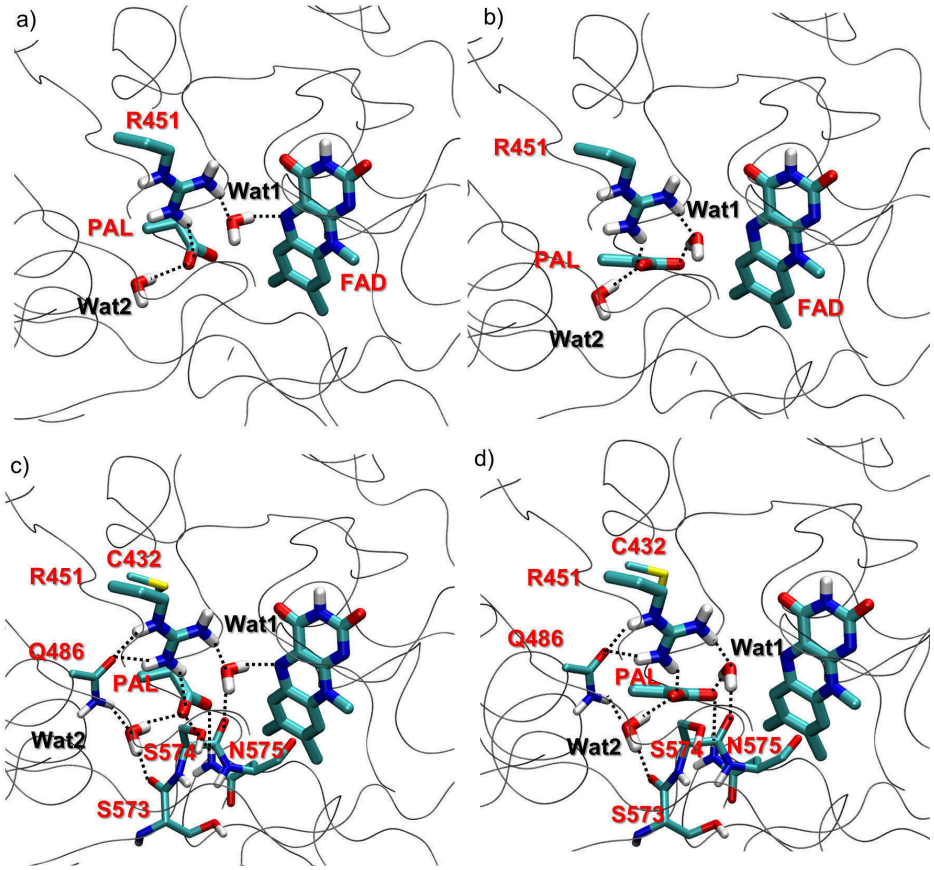


Figure 3.14: QM/MM optimized structure of the first CT state and LE state at the CAM-B3LYP/def2-TZVP level of theory. QM region (a) and active site (c) of the optimized CT state geometry; QM region (b) and active site (d) of the optimized LE state geometry.

In summary, the present calculations corroborate the hypothesis that the deprotonated form of palmitic acid ($RCOO^-$) is strongly favoured in the CvFAP active site, excluding the alternative mechanism proposed by Wu et al. [118]. More specifically, TD-DFT calculations within the QM/MM framework show that the charge-transfer (CT) state between FAD and PAL, which is fundamental for the reaction, lies at 4.15 eV in the palmitate configuration (RCOO-). In contrast, when palmitic acid is present (RCOOH), no CT state is found below 6.30 eV.

An additional evidence of the protonation state of PAL can be obtained from the calculation of the pKa of palmitic acid within the active site of the protein. The average pKa value calculated with PropKa tool [82] on 15 different structures, extracted from the MD simulation, is 4.79 and this is perfectly consistent with the standard value in aqueous solution (around 4.50-4.75). However, this result should be interpreted with caution, as pKa predictions for ligands within protein binding pockets can be affected by local dielectric effects not precisely reproduced by the PropKa tool. Despite these limitations, at the working pH of 8.5, the deprotonated form (palmitate) is expected to be predominant.

3.2.2 The simulation of FAD absorption spectrum in CvFAP

Once the protonation state of the substrate and the nature of charge-transfer states is defined by previous TD-DFT calculations, conducted at CAM-B3LYP/def2-TZVP level of theory, our focus is shifted to the simulation of the UV-Vis spectrum of FAD in CvFAP. In particular, our interest lies in understanding whether the red-shifted FAD peaks in the resting state is more related to the effect of the butterfly bending or to environmental contributions. Moreover, another important point is achieving a deeper comprehension regarding the formation of the more red-shifted FAD species (FAD_{RS}).

In this context, we use a hybrid quantum/classical approach, based on the MD-PMM method and QM/MM calculations. This approach allows extensive sampling of the enzyme's conformational space and reproduces the perturbative effects of the environment on the active site.

The choice of the DFT method here is not trivial and requires a careful assessment: range-separated DFT functionals are generally more accurate for the study of CT states, but are usually not very reliable in reproducing the position of experimental spectral bands. In recent benchmarks reported in the literature, conducted on a dataset of organic aromatic molecules, the two best DFT functionals (excluding double-hybrid functionals) for spectral calculations are B3LYP and M06 [125, 126].

However, these benchmarks are not specific to flavin compounds, so here benchmark calculations were performed in a QM/MM framework to assess the DFT functional that was able to better match the observations for this specific case. The benchmark, reported in figure S1, showed that B3LYP and M06 functional in conjunction with def2-TZVP basis set obtained the best results considering the two FAD's main absorption peaks positions and their relative intensities. Double-hybrid DFT functionals could not be used because, in the current ORCA implementation (ORCA 6), the calculation of excited states dipoles, that is crucial for MD-PMM calculations, is not yet supported for these functionals.

The benchmark was further validated through an additional test calculation, where the M06 and B3LYP functionals, in conjunction with the def2-TZVP basis set, were tested in a short MD-PMM calculation considering a 10 ns segment of the molecular dynamics trajectory of lumiflavin in water. As expected, the B3LYP spectrum exhibited closer agreement with the experimental data; therefore, this functional will be adopted for our MD-PMM calculations. The results of this preliminary MD-PMM calculation are reported in figure S2 of the SI.

Despite extensive benchmarking, the TD-DFT results were still far from matching the absolute positions of the observed spectra. Hence, in accordance with previous

MD-PMM studies [40, 32, 127], a scaling factor has been applied to the excitation energies. In this study, a value of 0.94 was selected, thereby imposing a match of the MD-PMM first calculated peak (418 nm) with the 445 nm first experimental peak in the lumiflavin absorption spectrum in water.

Quantifying the effect of Lumiflavin conformation on absorption spectra

This section addresses the effect of lumiflavin molecular bending on the observed FAD red-shift in the absorption spectrum, excluding contributions from Coulombic interactions, van der Waals forces, and environmental perturbations.

Initially, a molecular dynamics simulation of the *apo* protein was conducted for 100 ns using the CHARMM force field (further details are reported in the methods section). The specific choice between the *apo* and *holo* forms was deemed not important because experimental data indicate negligible differences in both peak positions and intensities between *holo* and *apo* (as reported in figure 3.11). By focusing on the apo-protein, we can draw more general conclusions.

From the MD trajectory, three representative lumiflavin conformations were extracted, corresponding to three values of the C_4–N_5–N_{10}–C_9 dihedral angle: bent (20°), partially bent (10°), and planar (0°) conformations. The three structures were then optimized at the B3LYP/def2-TZVP level of theory in gas-phase constraining the six dihedral angles involved in the "butterfly bending" motion (see methods section for further details). In this case, the focus is on accounting for the effect of the bending mode only, while excluding possible contributions from other conformational effects, such as isoalloxazine ring twisting. Therefore, following the definition provided by Schapiro et al. [128], all selected structures have a twisting angle (defined here as N_3-C_2-C_7-C_8) close to the planar value in order to exclude any possible contribution from twisting. This choice is consistent with molecular dynamics data, which indicate that the majority of the structures remain essentially untwisted. Gas-phase TD-DFT calculations

were then carried out at the same level of theory on these optimized geometries. Their corresponding absorption spectra are shown in Figure 3.15 and in Table 3.3 (See also tables S3-S5 for more details).

Our calculations indicate that the structural bending produces only a modest shift in peak positions, approximately 4 nm from the planar to the bent structure. The main effect is, instead, on the relative intensity of the two principal peaks, which increases from about 0.76 for the planar structure to 0.97 for the bent conformation. These results are noteworthy: despite the computed gas-phase peak positions being far from the experimental values (29 nm for the first band and 45 nm for the second) the predicted relative intensities for the weighted structure (0.85) closely match the experimental pattern (1.01).

This simple computational experiment suggests that the surrounding environment is expected to be predominant in determining the absolute positions of the absorption bands, whereas the bending influences the relative intensities. These "unperturbed data" will be refined in the subsequent MD–PMM calculations, explicitly introducing environmental effects.

Table 3.3: a) TD-DFT gas-phase positions of the main lumiflavin UV-Vis absorption peaks. and b) their relative intensity, for lumiflavin in various conformations at the B3LYP/def2-TZVP level of theory. Energies were corrected by a scaling factor of 0.94. Data were compared with experimental spectrum in CvFAP [118]

Geometry	λ_1 (nm)[a]	λ_2 (nm)[a]	I_{rel}[b]
Bent (20°)	441	347	0.97
Part. Bent (10°)	437	344	0.83
Planar	437	343	0.76
Weighted[c]	438	344	0.85
CvFAP *apo* [118]	467	389	1.01

[a] Position of the principal UV–Vis absorption peaks.

[b] I_{rel} refers to the relative intensity of the two main peaks ($I_{\lambda_1}/I_{\lambda_2}$).

[c] "Weighted" refers to the total spectrum obtained as a weighted average of the three lumiflavin conformations, where the weight assigned to each conformation corresponds to its population in the MD simulation of FAD in the *apo* state of CvFAP.

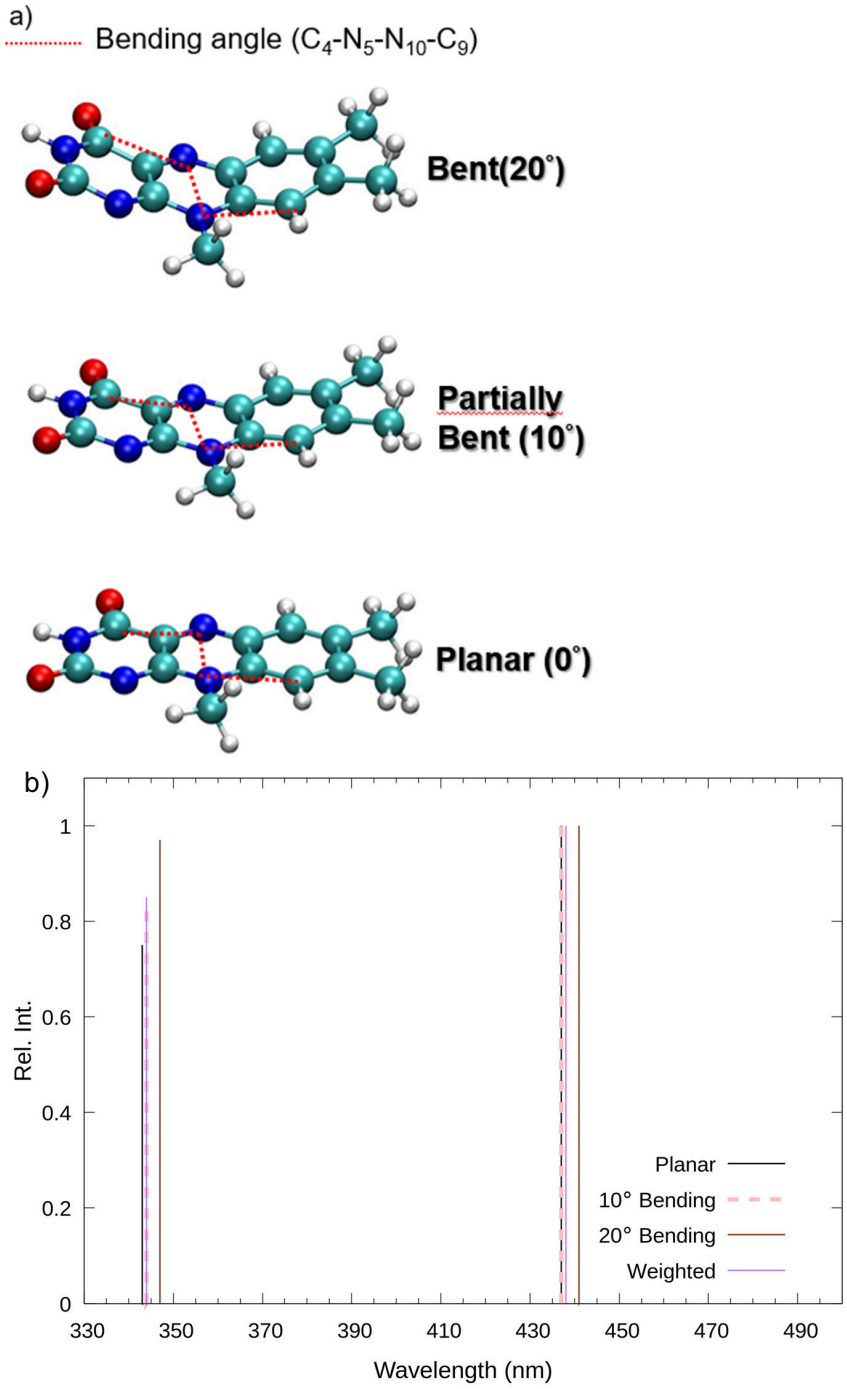


Figure 3.15: Correlation between the molecular conformation of lumiflavin and its absorption spectrum in gas-phase. a) Optimized conformations of lumiflavin, at the B3LYP/def2-TZVP level, obtained at different values of the C_4–N_5–N_{10}-C_9 dihedral angle (highlighted with red dashed lines): bent (20°), partially bent (10°), and planar (0°) structure. b) TD-DFT absorption peaks (nm) and normalized absorption intensities for the first two bright electronic states of the three conformers calculated in gas-phase at the same level of theory. A scaling factor of 0.94 was applied to the energies (in eV).

Effect of the molecular environment on the FAD Absorption Spectrum in CvFAP

To investigate the contribution of the molecular environment to the FAD absorption spectrum, MD-PMM calculations were performed. To assess how different environments could influence the spectra, two different MD simulations were conducted: one in the *apo* CvFAP and another in water. Additionally also the different FAD conformations were sampled considering the two MD trajectories. FAD remains mainly planar in the water simulation, instead, in the CvFAP trajectory, the bending angle varies significantly (see figure S3 for more details).

To account for both the environmental effects and the conformational variations in CvFAP, the corresponding MD frames were divided into three conformational subsets, according to the value of the bending angle of the lumiflavin moiety of FAD: planar ($\approx 0^\circ$ bending, 27% of frames), partially bent ($\approx 10^\circ$, 44% of frames), and bent ($\approx 20^\circ$, 29% of frames). For each subset, the corresponding unperturbed reference geometry (already reported in figure 3.15) and its associated eigenstates were used for the subsequent PMM calculations.

As shown in figure 3.16 and table 3.4 in all three basins, the inclusion of environmental perturbations by means of MD-PMM induces a bathochromic shift of the main absorption peaks by approximately 10-20 nm relative to the corresponding unperturbed values. Furthermore, the red-shift between the planar and 20° bent conformations remains essentially unchanged compared to the gas-phase spectra (4 nm).

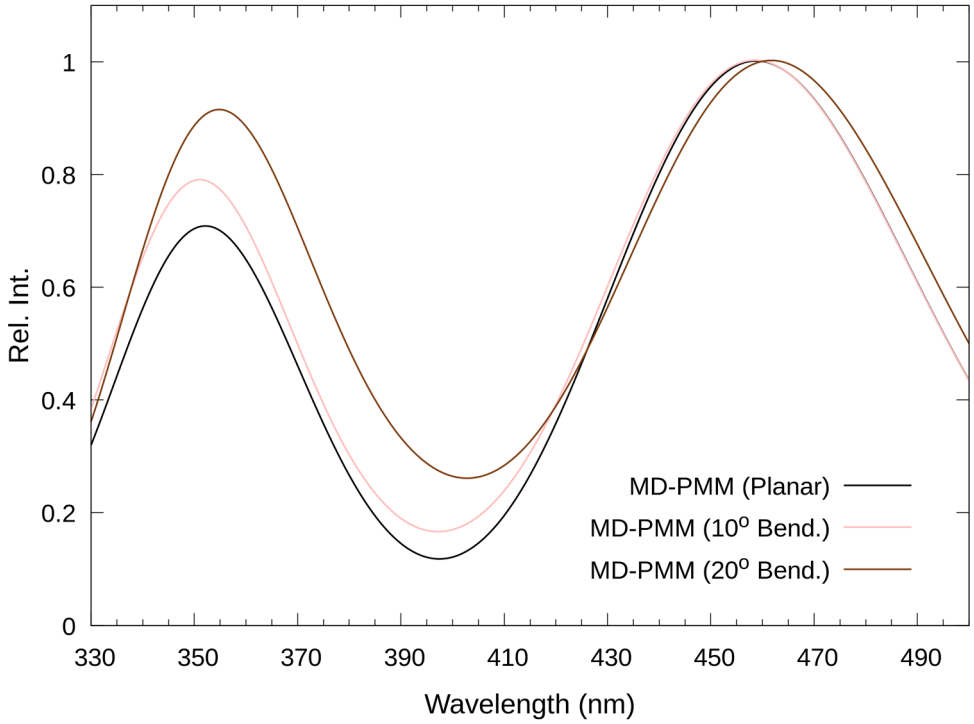


Figure 3.16: MD-PMM absorption spectra of the three conformations of FAD in CvFAP: planar (black), partially bent (pink), and bent (brown).

Table 3.4: MD-PMM positions (nm) of the main lumiflavin UV-Vis absorption peaks and their relative intensity (I_{rel}) for the three FAD conformations.

Conformation	λ_1(nm)^a	λ_2(nm)^a	I^b_{rel}
Planar	458	352	0.71
Partially bent	458	352	0.79
Bent	462	355	0.92

[a] Position of the principal UV–Vis absorption peaks, defined as the band center in the corresponding MD-PMM spectrum.
[b] Relative intensity of the two main peaks ($I_{\lambda_1}/I_{\lambda_2}$).

These data will be used to compute the "MD-weighted" CvFAP spectrum, obtained as a weighted average over the three conformations, where the weights correspond to the relative population (% of frames) of each conformational basin during the MD simulation. The resulting spectrum is compared with experimental spectra, and with the spectra in water (both computational and experimental). This comparison is reported in Figure 3.17 and in table 3.5.

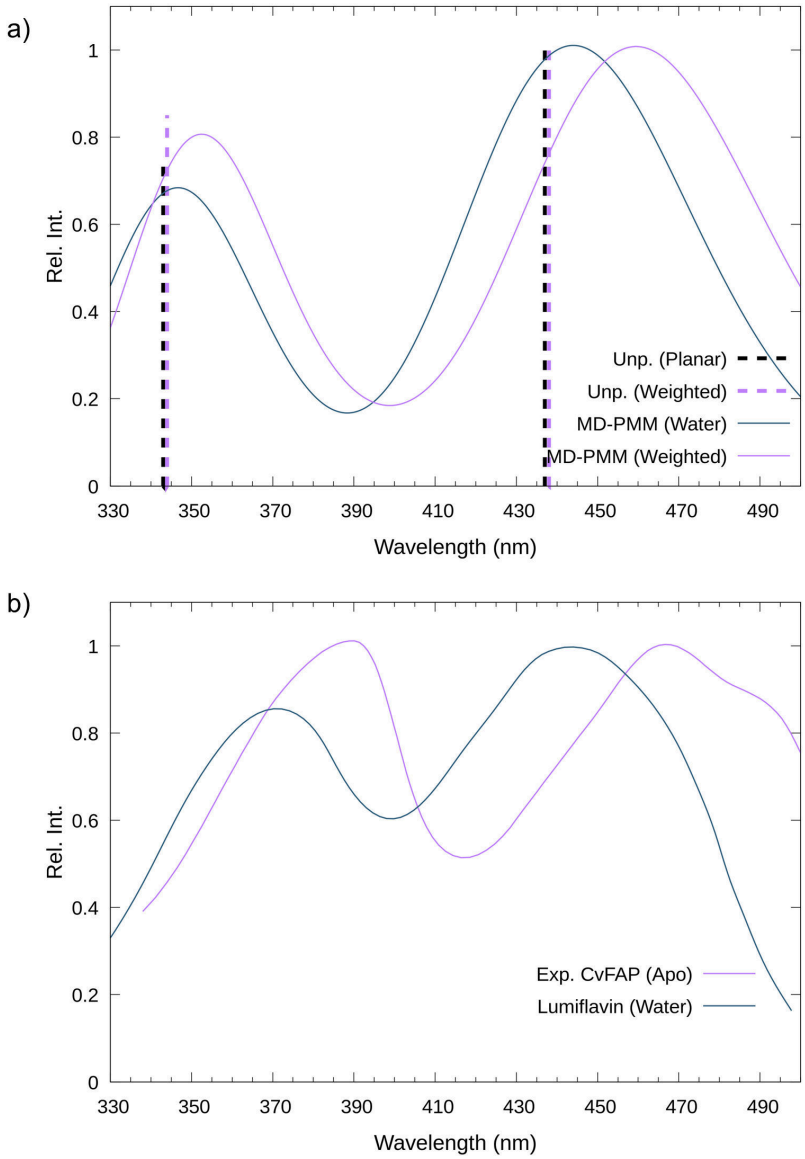


Figure 3.17: Comparison of MD-PMM (a) and experimental spectra (b) of lumiflavin in water and in CvFAP. The MD-weighted PMM spectrum in CvFAP (purple) closely matches the experimental one, reproducing both the relative peak intensity (0.81 vs. 1.01) and the red shift of the first absorption maximum (16 nm vs. 23 nm) compared to water (blue).

Table 3.5: Positions of the UV-Vis absorption peaks and relative intensities (I_{rel}) for FAD in CvFAP and lumiflavin in water, based on MD-PMM calculations and experimental data.

Conformation / System	λ_1 (nm)	λ_2 (nm)	I_{rel}
	MD-PMM		
Weighted (CvFAP)	459	353	0.81
LF in water	443	346	0.69
	Experimental		
CvFAP *apo* [118]	467	389	1.01
CvFAP *holo* [117]	467	391	0.96
LF in water [121]	443	370	0.86

The reported result highlights a qualitative agreement between the experimental and MD-PMM spectra, particularly regarding the position of the first peak (467 nm vs. 459 nm) and the relative intensity of the two peaks (0.81), which is close to the experimental value (1.01). Furthermore, MD-PMM reproduces the experimental spectral shift between water and the apoenzyme, amounting to approximately 20 nm.

An additional concept relevant to this thesis is that MD-PMM can accurately reproduce the enzyme's conformational variability, accounting for these dynamical effects in spectral calculations. Hence the QM/MM spectrum, obtained at the B3LYP/def2-TZVP level of theory, was compared to the MD-PMM calculated.

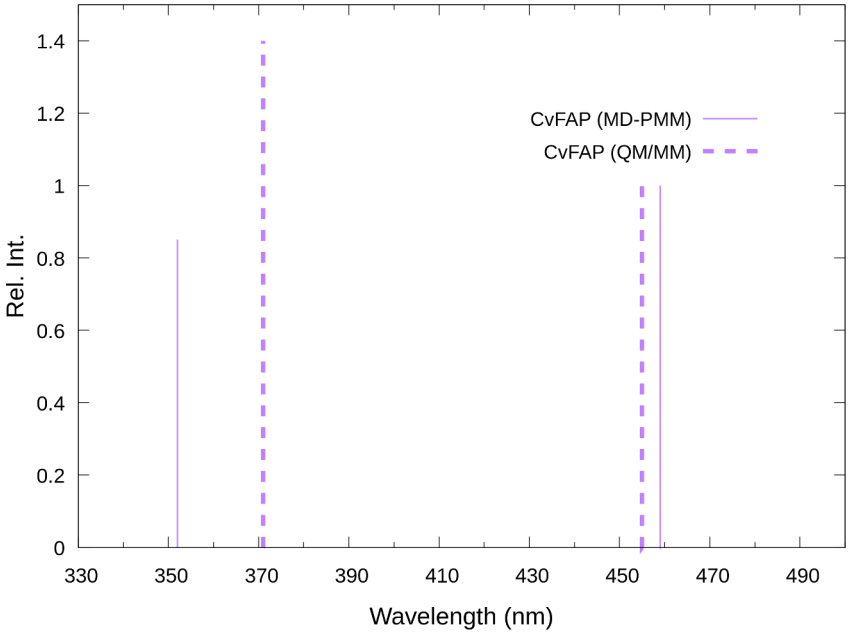


Figure 3.18: Comparison between the QM/MM and the MD-PMM spectra of FAD calculated in CvFAP. In both calculations only lumiflavin was included in the QM part. A scaling factor of 0.94 was applied to energies.

The results reported in Figure 3.18 show that MD-PMM is able to reproduce the relative peak intensities more accurately than the QM/MM calculation (0.80 vs. 1.4) when compared to the experimental value (1.01). Additionally, the first band in the QM/MM spectrum is blue-shifted by approximately 5 nm relative to the MD-PMM peak (459 nm), hence QM/MM shows a greater error compared to the experimental value (467 nm). These findings support the reliability of the MD–PMM methodology

for spectral calculations in complex biological systems.

Despite the promising outcomes, it remains essential to recognize the current limitations of the method. The accuracy of the MD-PMM calculations mainly depend on two factors: the level of theory of the unperturbed gas-phase calculations and the force field employed in the molecular dynamics simulations. In this specific case, the unperturbed gas-phase calculations were performed at TD-DFT level using the B3LYP functional and a triple-zeta basis set (def2-TZVP). Although this combination reproduces the experimental data reasonably well from a qualitative point of view, it is far from being exact, and therefore a scaling factor on the energies was introduced (unscaled calculations are reported in table S6). The second limitation of the method lies in the use of a non-polarizable force field (CHARMM36), since the current implementation of MD-PMM does not support polarizable force fields. However, the present results can be viewed as very promising even considering the limits of the methodology.

Future developments will be focused on the implementation of polarizable force fields within the PMM framework. In addition, also the inclusion of the vibrational contributions should be considered in further studies. These advancements will allow to achieve more accurate agreement with experimental data eliminating the need for a scaling factor.

Investigation of the more red-shifted intermediate

In this section the interest lies in understanding the origin of the more red-shifted FAD species observed in experimental studies [116, 117]. Upon illumination, the FAD main peak is decreased in intensity and red-shifted to around 480 nm with a shoulder appearing at 517 nm. These findings are considered to be associated with the formation of FAD_{RS} [117].

According to the mechanistic hypothesis already reported in 3.2.2 , after the $RCOO^-$ decarboxylation, there is a back-electron transfer (BET) from $FAD^{\bullet -}$ to $R^\bullet$, accompanied by a protonation step from an unidentified proton donor, completing the formation

of the alkane. This last step is associated with the formation of the aforementioned FAD_{RS} species.

Two different hypotheses have been proposed: according to Heyes et al. [117], the proton donor is Cys432; hence, the red shift could arise from a change in the protonation state of Cys432, from its neutral to its anionic form. Sorigue et al. [116], on the other hand, indicated Arg451 as the proton donor and supported the hypothesis that the red shift is linked to a change in the charge distribution of the active site: in the reagent structure there are two charged species: $RCOO^-$ and $Arg451^+$, in the intermediate state CO_2 and deprotonated Arg451. Both states are overall neutral, but in the reactant the charges are strongly localized, whereas in the intermediate state all species are neutral.

To verify this hypotesis, the FAD-weighted absorption spectrum was therefore computed again with the MD-PMM method using the same molecular dynamics trajectory, but with a modified PMM topology in which the total charges of Cys432 and Arg451 were altered. Two independent calculations were carried out: in the first, the total charge of Arg451 was reduced from +1 to 0 by scaling its side-chain charges.; in the second, the charges of Cys432 were scaled to yield a total net charge of –1. The resulting MD-PMM weighted spectra, reported in Figure 3.19 (see also figure S4 for more details), indicate that both modifications produce a blue shift of the absorption peaks, rather than the expected red shift. This strongly suggests that the formation of FAD_{RS} is not attributable to a purely electrostatic effect. We instead hypothesized that the red shift could originate from a geometric rearrangement of the active site upon transition from the reactant to the product structure, thereby motivating further QM/MM investigations. The starting reactant structure was optimized with QM/MM using the same protocol described in Section 3.2.2, but including also the side chain of Cys432 in the QM region. Nevertheless, the resulting optimized structure does not exhibit significant differences from that previously already reported in Figure 3.12.

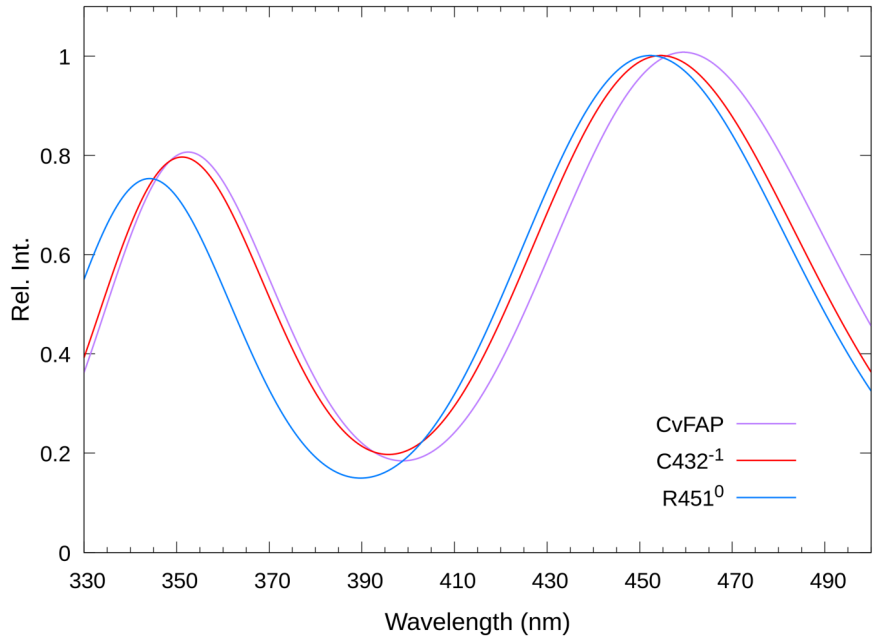


Figure 3.19: Comparison between the MD-weighted spectrum in CvFAP obtained using MD-PMM calculations with the standard charges and those with altered charges, namely neutral Arg451 and negatively charged (-1) Cys432. In both cases, the charge alterations result in a blue shift, rather than in a red shift, more pronounced for the neutral Arg451.

The product structure, was taken from the most favorable reaction pathway identified by Sorigué et al. [116] (denoted as Path II). Following decarboxylation, a proton-coupled electron transfer (PCET) occurs in which Arg451 protonates the substrate to form RH. Arg451 is then reprotonated by Wat1, and the resulting OH^- reacts with CO_2 to form bicarbonate in the final product state. Importantly, Sorigué et al. provided strong experimental evidence indicating that the mechanism suggested by Heyes et al. is unlikely, whereas Path II represents a more plausible pathway. Therefore, in this work, only the product corresponding to Path II was considered.

Optimized structures are reported in Figure 3.20. In the product structure, the palmitate head group and Wat1 were replaced by a bicarbonate ion and the pentadecane head group. The QM/MM-optimized structures reveal that neither the bending nor the twisting dihedrals of FAD undergo a significant change. However, in the product structure a hydrogen bond is present between FAD N5 and the bicarbonate ion.

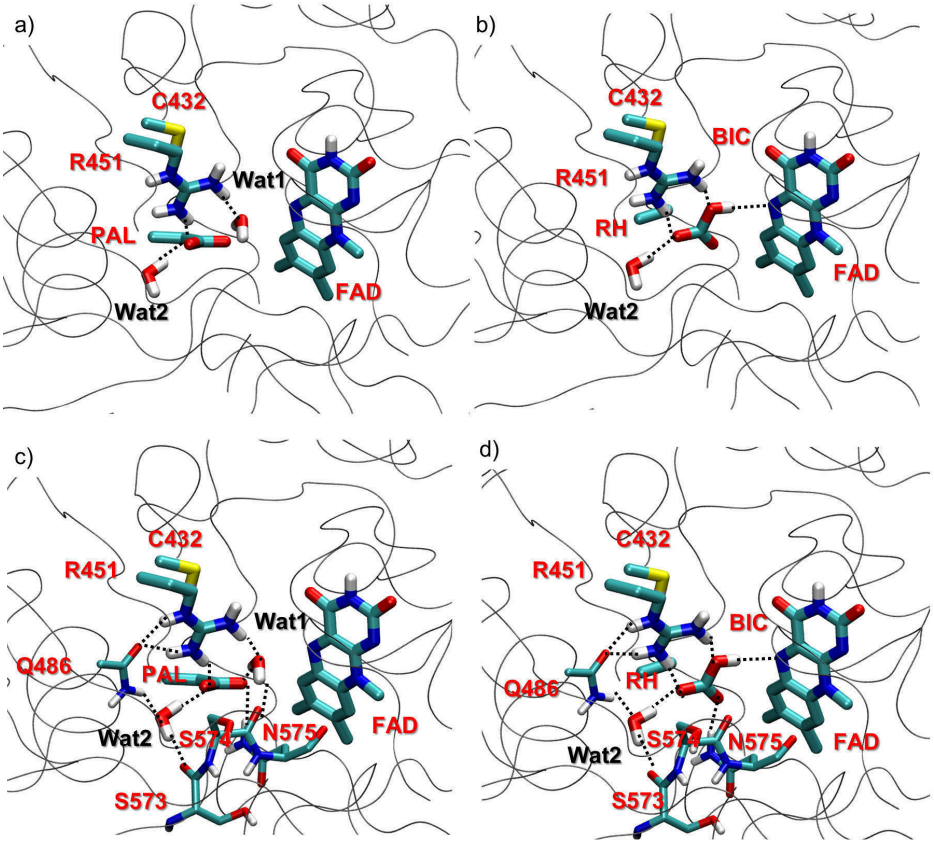


Figure 3.20: QM/MM-optimized structures for the reagent and product structures: (a) QM region of the reagent structure; (b) QM region of the product configuration; (c) reagent active site; (d) product active site. The QM region, in the reagent includes the head of the palmitate ion (-CH_2-CH_2-CO_2^- fragment), the lumiflavin moiety of FAD, two catalytic water molecules (Wat1 and Wat2) and the side chain of C432 and R451. In the product region, the bicarbonate ion and the alkane head RH (-CH_2-CH_3 fragment) replace the head of the palmitate ion and the Wat1. The active site representation includes also the MM residues Q486 (side chain) S573, S574, and N575.

After geometry optimization, the TD-DFT spectra of both the reactant and product structures were calculated at the same level of theory. The results, summarized in Figure 3.21, show that the QM/MM spectrum of the product structure is in good qualitative agreement with the experimental spectra recorded after illumination: the first peak is red-shifted to 478 nm and exhibits a lower intensity. However, our calculations could not reproduce the experimental tail observed at 517 nm, which is likely associated with vibrational contributions.

Moreover, the intensity of the second peak (390 nm) was not accurately reproduced by a single QM/MM calculation, which represents a well-known limitation of this method. Nonetheless, in this part of the study, the focus was on the first peak while the second

peak is less relevant. Despite this limitations, the present QM/MM results achieve a satisfactory qualitative agreement with the experimental data.

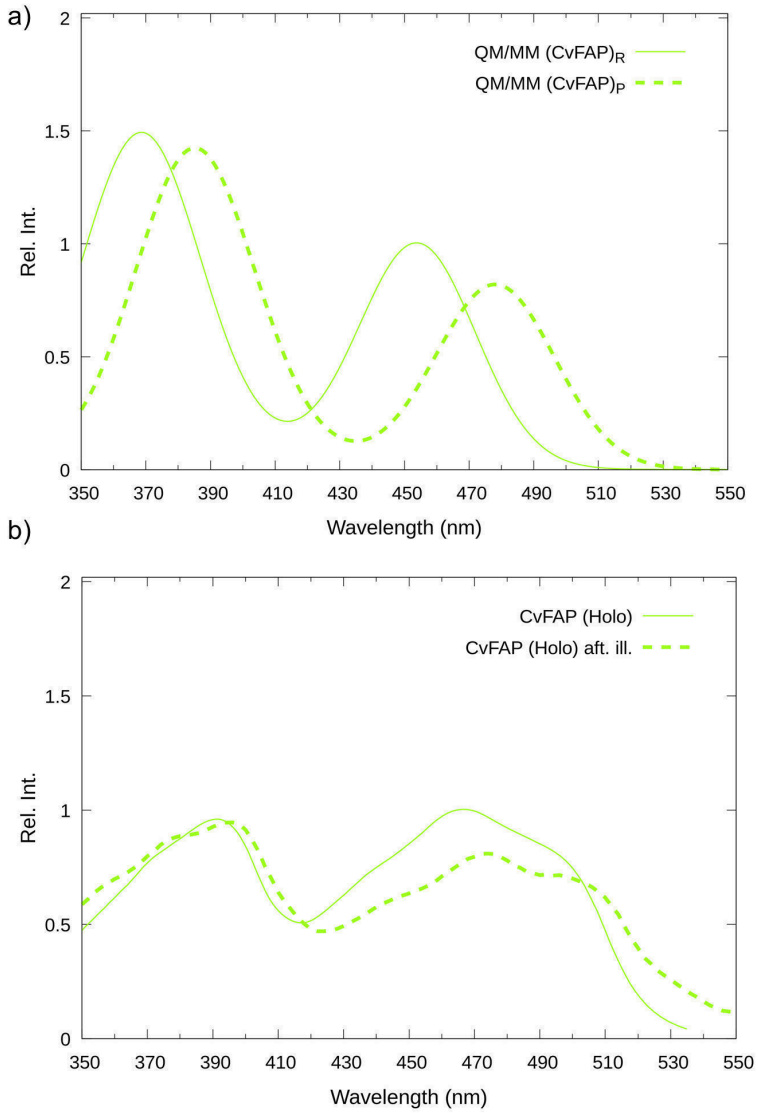


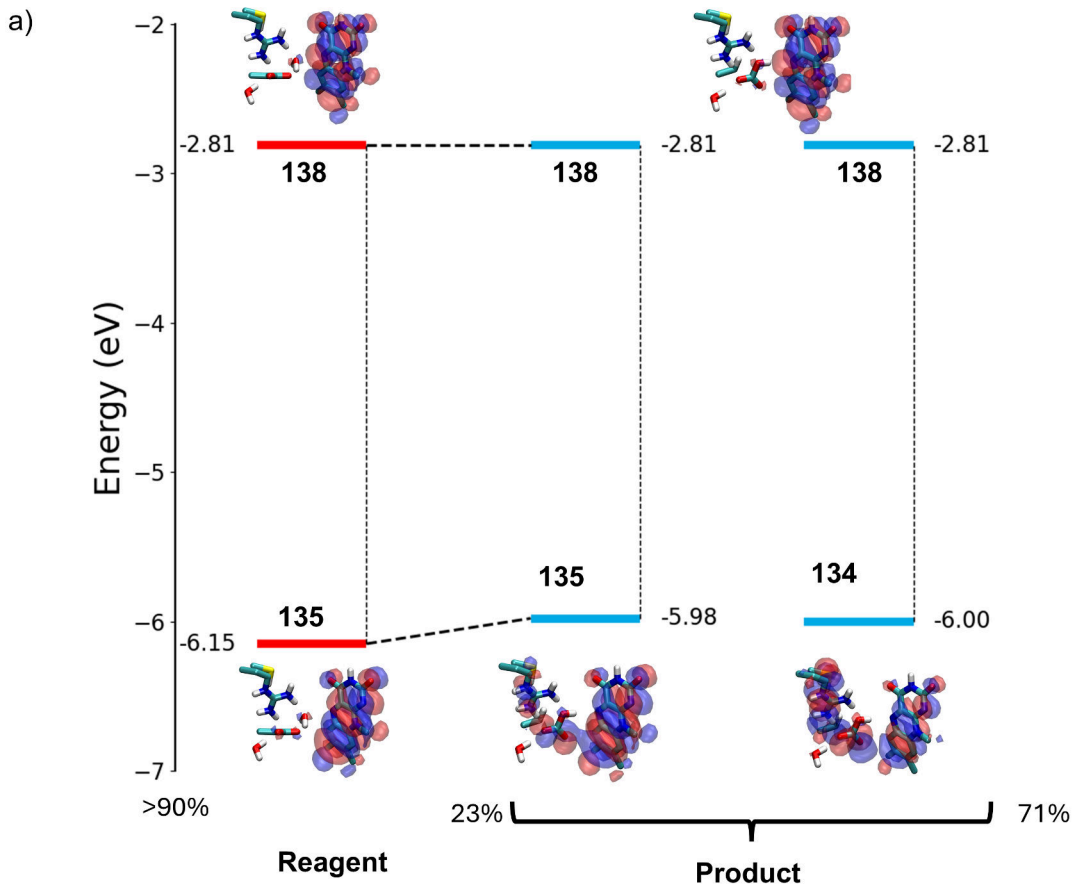
Figure 3.21: (a) Comparison of the FAD absorption peaks between the QM/MM spectra of the product (dashed line) and the reactant structure (solid line). (b) Experimental spectra of CvFAP [117] in the dark state (solid line) and after illumination (dashed line). Calculations were performed in a QM/MM setup at the B3LYP def2-TZVP level of theory. Atoms included in the QM are reported in figure 3.20. A scaling factor of 0.94 was applied on energies.

Our initial hypothesis was that the proximity between a negatively charged species (bicarbonate) and the N_5 atom of FAD could be the main cause of the red shift, due to polarization effects. In recent years, Kabir et al. [129] have shown that the presence of a charged residue near the N_5 or N_1 atoms of FMN can polarize N_1, N_5, and C_{4a} atoms of the isoalloxazine ring, inducing a significant change in the relative

absorption peaks. Hirshfeld charge analysis calculation was performed on both the reactant and product structures, and it surprisingly revealed no significant variation in the partial charges of N_1, N_5, and C_{4a} shifting from the reactant to the product structure, excluding a possible polarization of FAD related to hydrogen bond. Hence, our analysis focused on the molecular orbitals involved in the electronic transitions related to the first bright state.

The TD-DFT molecular orbitals show that, in the reactant, the transition is dominated by a single excitation from orbital 135 (Energy = $-6.15\,\mathrm{eV}$) to orbital 138 (Energy = $-2.81\,\mathrm{eV}$). In the product, the transition is mainly composed of two dominant contributions. The first, accounting for 71% of the transition, corresponds to a single excitation from orbital 134 ($E = -6.00\,\mathrm{eV}$) to orbital 138. The second, contributing 23%, arises from a single excitation from orbital 135 to 138. Interestingly, the orbital 138 remains at the same energy in both the reactant and the product, while orbitals 134 and 135 are higher in energy in the product. Specifically, orbital 135 is destabilized from -6.15 to $-5.98\,\mathrm{eV}$.

Additionally, in the reactant structure, orbital 135 is strongly localized on lumiflavin, while in the product, both orbitals 134 and 135 are delocalized over a significant portion of the reaction center, as reported in Figure 3.22. This reduced transition energy gap in the product structure is leading to the observed red shift. This energy difference is likely linked to a reorganization of the quantum center upon product formation. Additionally analysis of the NTOs involved in the same transition (Figure 3.22) highlights negligible changes in both the donor and acceptor orbitals, showing the lack of any variation in hole density.



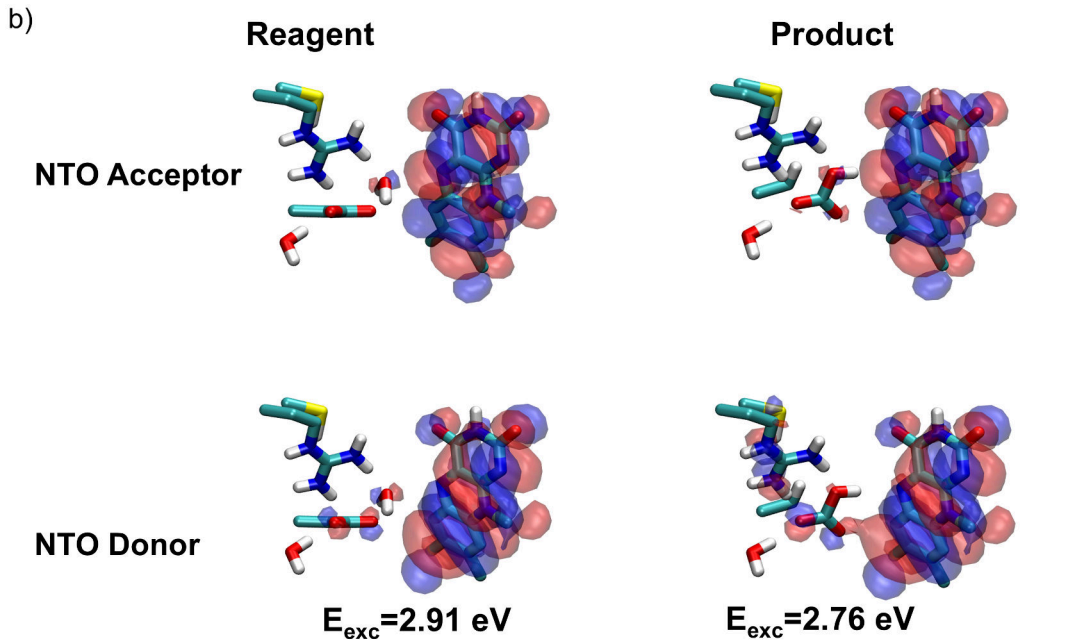


Figure 3.22: Characterization of the molecular orbitals involved in the first bright transition: a) Comparison of the first bright transition energy gap in reagent (red orbitals) and product structures (blue orbitals). The product shows a smaller gap due to destabilization of the initially occupied orbital during formation. Corresponding molecular orbitals are displayed (blue = negative, red = positive). b) NTOs related to the first bright transition in both structures.

Final Remarks

In conclusion, our results unambiguously demonstrate that the substrate is deprotonated within the active site, ruling out the alternative mechanism proposed by Wu et al. [118]. Moreover, our MD-PMM calculations indicate that the red shift of the FAD absorption peaks in the resting state arises primarily from environmental perturbations, while the FAD conformation (bending) mainly influences the relative peak intensities. Finally we investigated the origin of the transient FAD_{RS} species: our findings suggest that the formation of this species is unlikely to be caused by simple electrostatic interactions, but rather is more plausibly associated with a substantial reorganization of the active site linked with the decarboxylation process.

Chapter 4

Conclusions

The work presented in this thesis represents an important application of hybrid computational methods to the study of reaction mechanisms in complex systems. The primary focus of the entire work is on biomolecular systems and, in particular, on enzymes of considerable interest in the field of biocatalysis. These systems were selected as prototypical examples of complex catalytic environments, characterized by noticeable structural flexibility, large size, and reaction mechanisms that remain poorly understood. Also, their possible industrial relevance further justify their scientific interest.

Specifically, two techniques were employed: QM/MM schemes (i.e., ONIOM), used to gain insight into the reaction mechanisms, and MD–PMM, applied to quantitatively assess environmental and dynamical effects, in particular, to evaluate how the protein environment and its fluctuations can influence the reaction occurring within the enzymatic pocket. From a computational perspective, our work represents an elegant integration of QM/MM and MD-PMM approaches for mechanistic studies. The approach is of general applicability and is computationally less expensive compared to ab initio molecular dynamics simulations.

In the first system (HLADH) we studied the possible reaction mechanisms for the enantioselective reduction of 2-pentanone to *(S)*-2-pentanol in the presence of the

biomimetic co-factor BNAH. Our calculations have shown that the initial step of the reaction is a direct hydride transfer from C_4–H of the biomimetic co-factor BNAH to the carbonyl group of the substrate (2-pentanone), forming the *(S)*-pentan-2-olate. This step features an energy barrier of 14 ± 2.4 kcal mol^{-1} accurately estimated through MD-PMM calculations.

After this step, His51 is protonated by the bulk. Finally, the proton is released from His51$^+$ to the oxygen atom of the *(S)*-pentan-2-olate through a proton-transfer chain, yielding the *(S)*-alcohol product. In particular, His51$^+$ donates a proton to the oxygen atom of Ser48 through the mediation of one or more water molecules and subsequently the proton is transferred from Ser48 to the substrate. The presence of water molecules creating a hydrogen-bond network between His51 and Ser48 is crucial for this step.

At the same time, the calculations has shown that the alternative mechanism proposed by Marrone et al. [16], instead, is not viable since here the hydride transfer step has a prohibitively high Gibbs free energy barrier of around 44 kcal mol^{-1}.

This mechanistic hypothesis, if corroborated by further experimental evidence, could pave the way to the use of the biomimetic co-factor BNAH as an effective substitute for 1,4-NADH thererfore significantly lowering production costs for the synthesis of chiral *(S)*-alcohols and other enzyme-catalyzed products.

The second study, conducted on the CvFAP photoenzyme, is more complex and can be divided into two parts. The first is a preliminary phase aimed at determining the protonation state of the substrate (palmitic acid) within the active site. This investigation was necessary because, although many authors agree that the substrate is deprotonated before entering the active site [116, 97, 130, 117], a recent study [118] proposed an alternative reaction mechanism in which the substrate remains neutral in the catalytic site.

Since the CvFAP-catalyzed decarboxylation involves a charge-transfer (CT) state from the substrate (PAL) to the FAD cofactor, it was crucial to investigate the nature of the excited states. Our calculations revealed that a vertical PAL-to-FAD CT state exists

for the deprotonated conformation (RCOO-) at around 4.15 eV, whereas it is found at 6.31 eV in the system with palmitic acid (RCOOH). Considering these findings, it is therefore evident that the alternative mechanism proposed by Wu et al. [118] is unlikely. Therefore, in the subsequent spectroscopic study, only the deprotonated form of the substrate was considered.

The second part of the study focused on the resting-state absorption spectrum of FAD in CvFAP. This spectrum is particularly relevant because, among the many flavoproteins incorporating FAD as a cofactor, CvFAP exhibits the most strongly red-shifted FAD spectral peaks. Additionally, in this system, the lumiflavin moiety of FAD is bent in the oxidized form, a feature that is generally uncommon in proteins.

The main focus was to determine whether the dominant contribution to the observed red shift mainly comes from the conformation of FAD, specifically the bending of the lumiflavin ring, or from the perturbation exerted by the protein environment. MD simulations were used as the source of representative structures differing in the bending angle. Subsequently, gas phase TD-DFT and MD-PMM calculations were performed on these structures.

Our results clearly underline that, although bending has a significant effect on the relative intensity of the peaks, the main contribution to the observed red shift arises from the perturbation exerted by the protein environment. Consequently, the MD-PMM approach turned out to be important to account for the effect of environmental fluctuations on the absorption spectrum. Nonetheless, it should be emphasized that, from a conformational point of view, this work specifically addresses only the effect of the FAD butterfly bending, as it is currently defined in the literature [116]. However it is not excluded that other conformational variations, especially the isoalloxazine ring twisting, may influence the absorption spectra. These aspects are currently being investigated in a dedicated study that is still in its early phase.

Our investigation then focused on the origin of a transient FAD species, denoted FAD_{RS}, which arises after fatty-acid decarboxylation and exhibits an even more pro-

nounced red shift in its UV–Vis spectrum. In this context, QM/MM calculations were employed to elucidate the electronic factors underlying this additional spectral shift. A detailed analysis of the molecular orbitals involved in the first bright transition suggests that the additional red shift originates from a reorganization of the reaction center following the decarboxylation process. This reorganization involves the formation of a strong interaction between a bicarbonate anion and the FAD, leading to a change in the energy gap between the two molecular orbitals responsible for the transition. Our calculations provided a plausible explanation for the formation of FAD_{RS}.

These findings, also here, were made possible only by the integration of MD-PMM and QM/MM methodologies. In the HLADH case, the coupling was sequential, with MD-PMM calculations performed on structures previously optimized through QM/MM. In the CvFAP study, instead, we adopted a parallel coupling strategy, where the two sets of calculations were conducted independently. Nevertheless, the complementary insights provided by both approaches are essential for achieving a comprehensive understanding of the absorption spectrum.

Moreover, this comparison highlights the complementary strengths and limitations of the two methods. MD-PMM is excellent in capturing dynamical effects and reproducing experimental absorption peak positions and intensities with high accuracy. However, it cannot provide direct information about orbital interactions or about energy density distribution. In contrast, QM/MM enables a detailed analysis of frontier orbitals and structural rearrangements, but it neglects dynamical environmental fluctuations, often resulting in spectra with distorted relative intensities.

In conclusion, as illustrated in Figure 4.1, this work demonstrates that by integrating both computational methodologies it possible to achieve a thorough and precise understanding of complex reaction mechanism, including those underlying natural enzymatic processes.

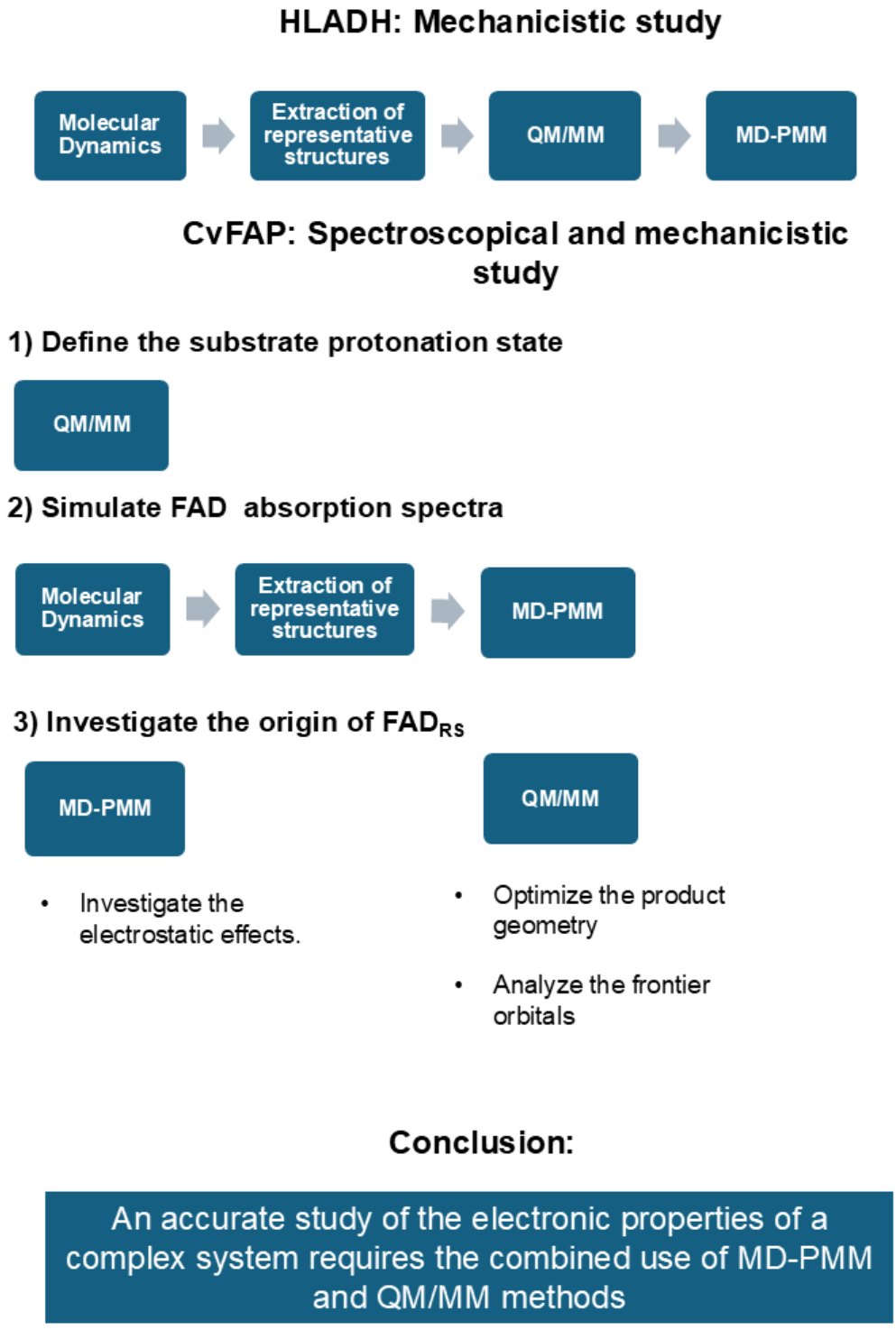


Figure 4.1: Summary of the computational workflow employed in this thesis.

Chapter 5

Appendix

5.1 Simulation of Raman Spectra of Dicationic, Imidazolium-Based, Ionic Liquids

Here, we report an additional study carried out during my Ph.D. that entails a direct collaboration with experimental research group of Dr.ssa F. Rondino (ENEA).

In particular, this work focused on the investigation of the Raman spectra of a series of imidazole-based dicationic ionic liquids. Conducted during the early stages of my Ph.D., it was an opportunity to gain practical experience with molecular dynamics (MD) software GROMACS and served as a "test study" for the methodological approach developed throughout my doctoral research.

Ionic liquids (ILs) are highly attractive systems due to their physicochemical properties: low vapor pressure, high chemical stability, high conductivity, and low melting point, which make them promising, greener alternatives to commonly used organic solvents [131, 132]. Typically, ionic liquids consist of a 1:1 combination of a charged cation and a charged anion; however, in recent years a special class containing a 2^+

charged cation (dication) coupled with two anions has been reported [133].

In this work, we aimed to reproduce the experimental Raman spectra (measured by our collaborators at the ENEA, centro ricerche Casaccia) of a specific series of dicationic ionic liquids, denoted as $[C_n(mim)_2](TFSI)_2$. As reported in figure 5.1, in these systems, the cation contains two 3-methylimidazolium rings linked by an alkyl chain $(CH_2)_n$, and this dication is paired with two bis(trifluoromethanesulfonyl)imide (TFSI) anions. The alkyl-chain length n was varied from 1 to 4. For simplicity we will denote the cations $[C_n(mim)_2]$ simply as C_n.

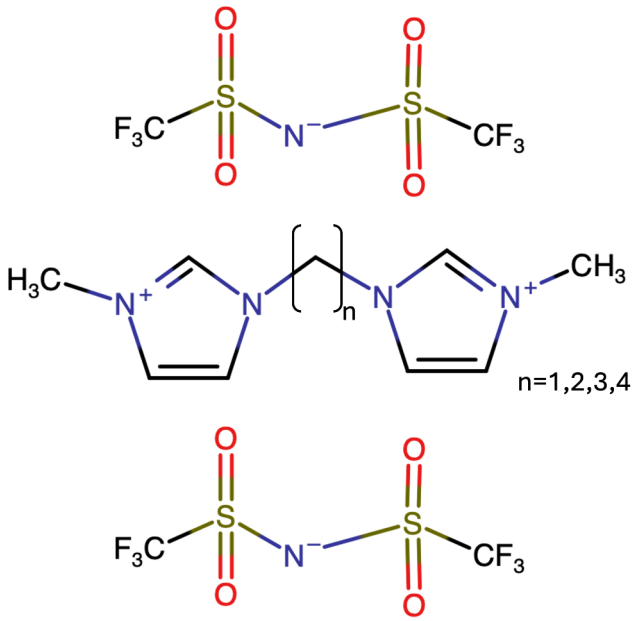


Figure 5.1: General structure of the dicationic ionic liquids studied in the work. Here n represent the number of C atoms in the alkyl chain linking the two 3-methylimidaziolium rings.

These systems are interesting because their chemical and physical properties can be tuned by varying the length of the alkyl linker between the two charged methylimidazolium rings. From a computational point of view, however, they pose a substantial challenge, as the liquid phase contains a large number of strongly interacting ions. Consequently, accurately reproducing the experimental Raman spectra requires to sample a large ensemble of molecular geometries. A further difficulty is that reliable Raman calculations generally demand large basis sets, which significantly increase

computational cost [134].

To address these issues, we adopted a combined strategy: firstly molecular dynamics simulations of these four systems were performed and then the Raman spectra was computed at the DFT level for a set of representative structures extracted from those trajectories. This workflow balances the need for wide conformational sampling with the accuracy required for reproducing the vibrational spectra.

5.1.1 Results and Discussion

The ionic liquid ensemble was decomposed into smaller structural units, which we define here as neutral triplet clusters (NTCs). Each NTC is composed of two anions and one cation.

Within each neutral triplet cluster, there are two conformational degrees of freedom: the intermolecular conformation, which is related to the arrangement of anions and cation within the cluster, and the intramolecular conformation, which represents the specific conformer adopted by each anion or cation.

Initially, to sample the possible intermolecular conformations, a molecular dynamics simulation of about 40 ns was performed for each ($[C_n(mim)_2](TFSI)_2$) ionic liquid. The system consisted of a box with 100 NTCs (200 anions and 100 cations). The simulations were conducted using the GROMACS software with the CL&P force field [135, 136]. The charges and missing parameters required to simulate these specific ionic liquids were adapted from previous studies [137, 138].

After the molecular dynamics run, tools based on essential dynamics principles [139, 140, 141] were used for conformational analysis. These Fortran tools allowed to divide the entire trajectory into different conformational basins and to identify the one corresponding to the most probable conformations, obtaining a subset of structures representing the most probable NTCs.

At this point, it was needed to investigate also the intramolecular conformation. Con-

sidering previous studies reported in the literature [142, 143, 144], to determine the exact structures of these conformers, we computed the dihedral distribution function of two significant dihedral angles, highlighted in Fig. 5.2: C–S–S'–C' for the anion and N–C_1–C_1'–N' for the cation. The results, shown in Fig. 5.2, indicate that the anions tend to adopt mainly two conformations: the *cisoid* (dihedral angle ±45°) conformation, where both CF_3 substituents are located on the same side relative to the main molecular axis, and the *transoid* conformation, with CF_3 groups on opposite sides (dihedral angle ±180°).

For the cations, the situation is similar: we refer to the *syn* conformation when the two rings lie on the same side with respect to the molecular axis defined by the alkyl chain, and *anti* when they are on opposite sides. In the case of C1, the dihedral does not exist and therefore only a single conformation is possible, whereas in the case of C2, although the dihedral exists, only the *anti* conformer was observed. The two longer-chain cations, namely C3 and C4, display both conformations. More specifically, in C3 the *anti* conformation is predominant with a dihedral angle of ±120°, and the minor *syn* conformation displays a dihedral angle of about 0°. In C4, by contrast, the *anti* configuration shows a dihedral angle of around ±180°, while the *syn* shows a dihedral angle of about ±60° and is the prevalent one. The simulation of Raman spectra was conducted for four systems. Starting from the previously identified basins containing the most probable NTC structures, from each basin, one representative structure was extracted for each conformer. Subsequently, we attempted to optimize these NTCs structures and to reproduce the Raman spectrum, considering both conformations and intermolecular interactions, in order to compare it with the experimental one. However this was not possible due to technical limitations: using a small basis set, the spectrum did not reproduce the experimental one at all, whereas with an appropriate basis set (Sadlej-PVTZ) [134] the computational cost was too high, with extremely prohibitive run times.

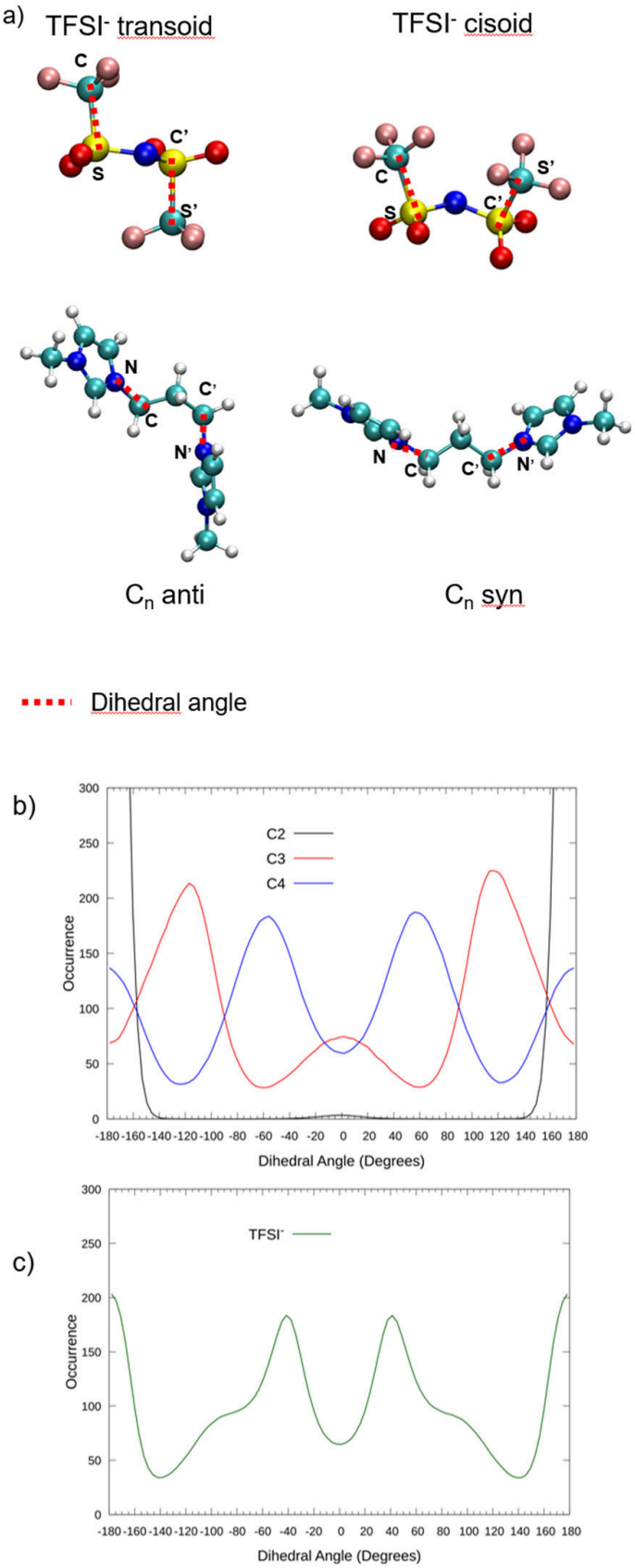


Figure 5.2: a) Structure of different intramolecular conformers for both cation and anion. b) Dihedral distribution functions for the three cations: C2, C3 and C4. c) Dihedral distribution function for the anion $TFSI^-$. The dihedral distribution function of the anion does not show any significant variations in all four systems.

In agreement with the approach of Lassegues et al. [145], who observed that for similar ionic liquids intermolecular interactions have very little influence on the Raman spectrum, the structures of the isolated cation and anion were extracted, taking into account the different conformers, and then the spectra were calculated as sum of isolated components. In particular, the structures of the individual cations and anions were extracted from the NTCs, considering for each the two possible conformations (syn/anti for the cation and cisoid/transoid for the anion respectively) and optimized at the wB97X-D3/Sadlej-PVTZ [146, 134, 147] level of theory. The structure is preserved, and a local minimum representative of that specific conformer is found very rapidly, without the need to impose constraints. This confirms the validity of the procedure used to extract the structures.

Here our calculated spectra are compared with the experimental ones.

The low-frequency region (260–440 cm^{-1}), it can be readily seen that it is dominated by the signals of the anion. This region, examined in Figure 5.3, shows that this zone is diagnostic for the two conformations of the anion but some interesting feautres of the cation also are present. As an example, the spectrum of $[C3(TFSI)_2]$ is reported, but the situation is similar also for C4.

The peak at 280 cm^{-1} is related to a coupled mode of the anion. The most interesting peaks are the ones associated with SO_2 rocking that are diagnostic for the anion conformation: the peak is present at 314 cm^{-1} for the transoid conformation and at 325 cm^{-1} for the cisoid conformation. The peaks related to SO_2 twisting/SNS bending also are, respectively, at 341 cm^{-1} for the transoid conformer and at 350 cm^{-1} in the cisoid conformer. With regard to the cation, it is worth focusing on a peak that emerges at about 300 cm^{-1} related to coupled C-N and C-H bending modes and it is diagnostic of the syn conformation of C3.

Finally, the signal at 400 cm^{-1} has both a contribution from the anion (SO_2 wagging) and one related to the C-H and C-N bending of cation.

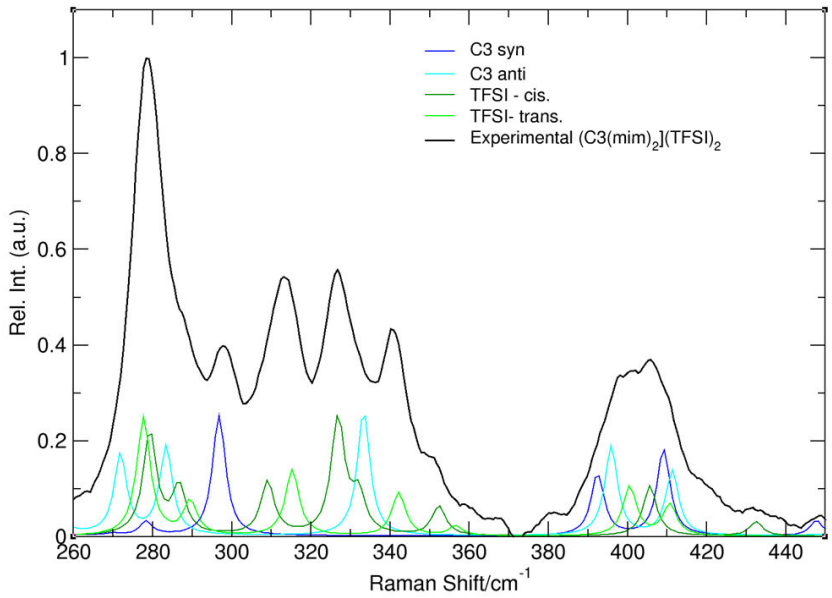


Figure 5.3: Comparison between simulated and experimental spectra of $[C3(mim)_2](TFSI)_2$ in the low frequency zones (260-440cm^{-1}). Peaks related to different conformers are highlighted. The applied scaling factor is 1.04 for the anion and 0.97 for the cation.

Moving further along the spectrum, the effect of the anion conformation is negligible; therefore, its average spectrum is reported. By analyzing the spectra of all systems, shown in Figure 5.4, the agreement between computational spectra and the experimental ones is very good. The main signal is related to the S-N-S stretching of the anion around 750 cm^{-1}. Other peaks related to the anion are found at 1132 (SO_2 symm. stretching / C-S stretching), 1200–1220 cm^{-1} (mainly C-F stretching), and at 1344 cm^{-1} (SO_2 asymm. stretching). Other bands in that zone are mainly attributable to the cation, with some differences between the two conformers found in the region around 600 cm^{-1}.The peak observed at 1020 cm^{-1} corresponds to the C-C stretching mode of the imidazole rings. The intense band at 1340 cm^{-1} is attributed to a ring breathing mode coupled with the CH_2 bending. Additional significant signals are found at 1400, 1419, and 1430 cm^{-1}, which are assigned to the stretching modes of the ring coupled with the CH_2 and CH_3 bending. In shorter systems (C1-C2), the last of these three bands appears more intense. Furthermore, a weak signal near 1450 cm^{-1} is associated with the CH_3NHCH bending mode of the ring. Finally, the region of 1500–1600 cm^{-1} is mainly related to the out-of-plane (o.o.p.) bending of C-H bonds.

Complete assignments are reported in tables 5.1 and 5.2.

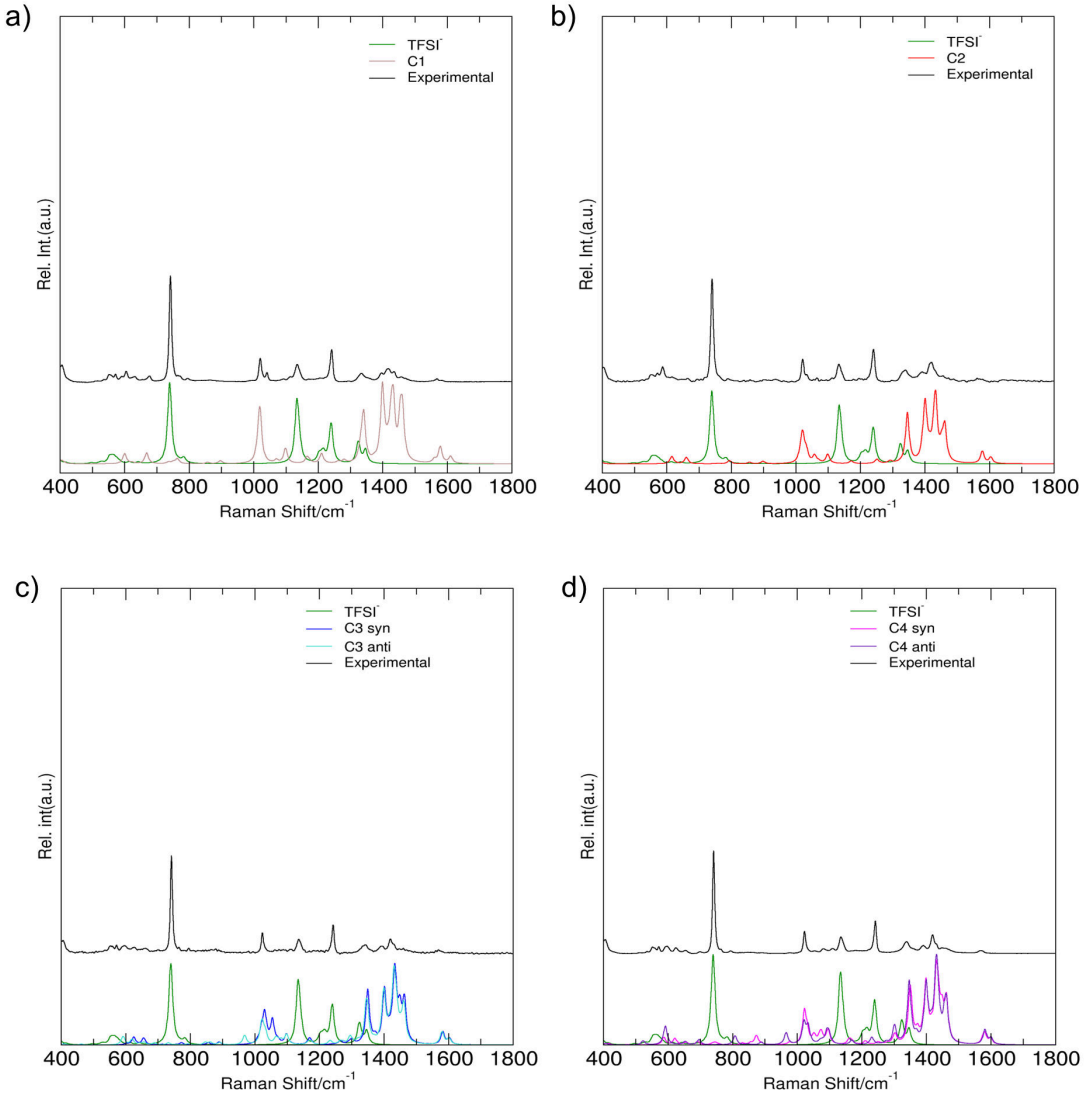


Figure 5.4: Comparison between simulated and experimental spectra of (a) $[C1(mim)_2](TFSI)_2$, (b) $[C2(mim)_2](TFSI)_2$, (c) $[C3(mim)_2](TFSI)_2$ and (d) $[C4(mim)_2](TFSI)_2$ in the fingerprint zone (400-1800cm^{-1}). Peaks related to different conformers are highlighted. The applied scaling factor is 1.04 for the anion and 0.97 for the cation.

Table 5.1: Assignment of the vibrational modes of the anion TFSI.

Raman Shift (cm-1)	Conformer	Normal Mode
280	Both	SO_2 bending/ CF_3 bending
314	Transoid	SO_2 rocking
325	Cisoid	SO_2 rocking
341	Transoid	SO_2 twisting/SNS bending
350	Cisoid	SO_2 twisting/SNS bending
400	Both	SO_2 wagging
500	Both	SO_2 rocking
519	Cisoid	SO_2 rocking / CF_3 rocking
529	Transoid	CF_3 rocking
560	Both	CF_3 rocking
580	Cisoid	C-S Stretching / C-F Symm. Stretch
613	Transoid	SO_2 bending
642	Cisoid	*(S)*-N-S symm. stretching
740	Both	S-N-S symm. stretching
781	Both	S-N-S symm. stretching
1132	Both	SO_2 symm. stretching / C-S stretching
1200	Both	C-F asymm. stretching
1217	Both	C-F asymm. stretching
1240	Both	SO_2/C-F symm. stretching
1322	Both	SO_2 asymm. stretching
1344	Both	SO_2 asymm. stretching

Table 5.2: Assignment for the main vibrational modes of cations Cn (n=1-4) in the fingerprint zone.

Raman Shift (cm-1)	Normal Mode
580-700	Strongly coupled modes:(C-N bending/stretching, CH_2 rocking)
750-900	C-H bending o.o.p
1020	C-C stretching
1030-1200	Very coupled bending modes
1300-1500	CH_2 bending/ Ring stretching/CH_3 bending
1570-1600	C-H o.o.p. bending

Finally, in the high frequency zone, reported in figure 5.5, only the signals of the cations are present. The bands are related to different stretching modes. The complete assignments, for each system, are reported in figure 5.1.

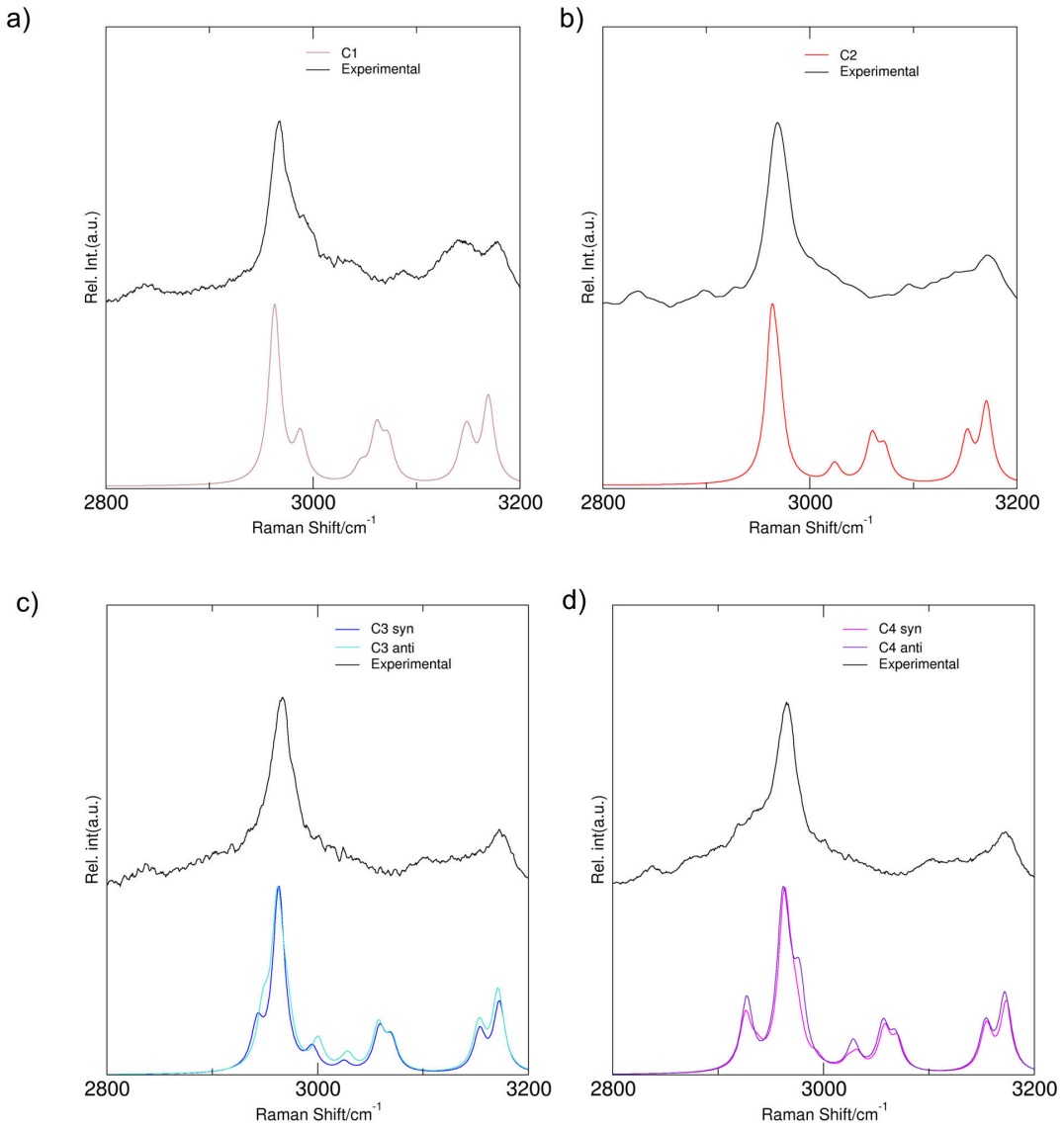


Figure 5.5: Comparison between simulated and experimental spectra of: (a) $[C1(mim)_2](TFSI)_2$, (b) $[C2(mim)_2](TFSI)_2$, (c) $[C3(mim)_2](TFSI)_2$ and (d) $[C4(mim)_2](TFSI)_2$ in the high frequency zone (2800-3200cm^{-1}). The applied scaling factor is 0.96 for the cation. No peaks related to $TFSI^-$ are present in that zone.

Table 5.1: Detailed assignment of the normal vibrational modes of cations Cn (n=1-4), for the high frequency zone.

Raman shift (cm^{-1})						**Normal Mode**
C1	C2	C3		C4		
		syn	*anti*	*syn*	*anti*	
/	/	2943	2948	2926	2926	CH_2 symm. stretch.
2960	2962	2963	2962	2962	2961	CH_3 symm. stretch.
2987	2970	2965	2971	2970	2977	CH_2 symm. stretch.
/	/	2995	3000	2993	/	CH_2 asymm. stretch.
3046	3024	3025	3022	3024	3028	CH_2 asymm. stretch.
/	/	/	3028	3032	/	CH_2 asymm. stretch.
3061	3060	3059	3058	3058	3056	CH_3 asymm. stretch.
3071	3071	3070	3069	3070	3068	CH_3 asymm. stretch.
3150	3151	3153	3153	3153	3152	C-H stretch.
3170	3170	3172	3172	3173	3172	C-H stretch.

5.1.2 Conclusions

Here, we have presented a solid computational approach for the simulation of Raman spectra of ionic liquids. Molecular Dynamics (MD) was used to account for the large conformational variability of the ions, while QM calculations were conducted on isolated species extracted from the MD trajectory. Hence, the total spectrum was obtained as the sum of the spectra of the isolated species (cation and anion), accounting for the different intramolecular conformers. The obtained results are in good agreement with experimental data; however, it is worth noting that this procedure can be applied only to ionic liquids where intermolecular interactions are weak, such as imidazolium-based ionic liquids, while its applicability could be limited when dealing with more complex ionic liquids.

5.1.3 Materials and method

Here only computational methods are reported. Further details about the experimental procedure are reported in the original paper [133]

Molecular Dynamics

All simulations were carried out with the GROMACS package [83], employing a refined version of the CL&P (Canongia-Lopes and Pádua) Force Field [135, 136], in which the partial atomic charges of the alkyl chain were adapted from previous works [137, 138].

This modified force field was validated by comparing the simulated densities with experimental values, showing good agreement.

The systems consisted of a cubic simulation box containing 200 anions and 100 cations.

Energy minimization was done with the steepest descent algorithm [148]. Molecular dynamics simulations were conducted with a timestep of 2 fs, applying a V-rescale thermostat (time constant of 0.1 ps) [88] and a Berendsen barostat [89] (τ= 5 ps, compressibility of 4.5×10^{-5}, P= 1 bar). All bonds with hydrogen atoms were constrained with the LINCS algorithm [87]. A cut-off of 1.2 nm was set for both van der Waals and Coulomb interactions, while long-range electrostatics were treated with the PME (Particle Mesh Ewald) method [86].

Equilibration was performed over 10 ns to ensure a stable density. Subsequently, a 40 ns production run was carried out using the Parrinello–Rahman barostat [99] ((τ= 15 ps, compressibility of 4.5×10^{-5}, P= 1 bar) together with the V-rescale thermostat (τ= 0.1 ps).

DFT calculations

Structures of neutral triplet clusters (NTCs), extracted from molecular dynamics simulations, were optimized at the B3LYP-D3/6-311G* level of theory. From these optimized NTC geometries, isolated cations and anions were obtained and further optimized using the wB97XD functional [146] in combination with the Sadlej-PVTZ basis set [134], which is specifically designed for Raman calculations. Subsequently, vibrational frequency calculations were performed at the same level of theory within the harmonic approximation. All quantum chemical calculations were carried out using the Gaussian 16 software package [149]. Finally, the Raman spectra were computed by averaging over the isolated species, considering different intramolecular conformations: syn and anti for the cation and cis and trans for the anion. This approach was in line with the one previously adopted by Lassegues et al. [145, 150]. The applied scaling factors were: 1.04 (250–450 cm^{-1} zone) and 1.00 (450–3200 cm^{-1} zone) for the $TFSI^-$ anion, and 0.97 (250–2800 cm^{-1} zone) and 0.96 (2800–3200 cm^{-1} zone) for the corresponding cation.

5.2 Supporting Information

5.2.1 CvFAP

QM/MM TD-DFT calculations

Table S1: Energy, oscillator strength (f.osc), wavelength, and nature of the first ten excited states of the QM region calculated at the CAM-B3LYP/def2-TZVP level of theory for the RCOOH configuration. In bold it is reported also the CT (PAL–FAD) that is the 19^{th} state.

Ex. State	Energy (eV)	Wavelength (nm)	f.osc	Nature
1	2.97	417.5	0.2156	LE(FAD)
2	3.67	338.0	0.0144	LE(FAD)
3	3.82	324.6	0.1664	LE(FAD)
4	4.12	301.9	0.0012	LE(FAD)
5	4.71	263.3	0.0145	LE(FAD)
6	4.83	256.6	0.0009	LE(FAD)
7	5.00	248.0	0.0663	LE(FAD)
8	5.09	243.6	0.0156	LE(FAD)
9	5.15	240.8	0.0144	CT(ARG–FAD)
10	5.17	239.7	0.5759	LE(FAD)
...	...	...	...	...
19	**6.31**	**196.6**	**0.0001**	**CT (PAL–FAD)**

Table S2: Energy, oscillator strength (f.osc), wavelength, and nature of the first ten excited states of the QM region calculated at the CAM-B3LYP/def2-TZVP level of theory for the $RCOO^-$ configuration. In bold it is reported the CT (PAL–FAD) that is the 4^{th} state.

Ex. State	Energy (eV)	Wavelength (nm)	f.osc	Nature
1	3.22	385.1	0.2370	LE(FAD)
2	3.65	339.7	0.0048	LE(FAD)
3	3.94	314.5	0.1791	LE(FAD)
4	**4.15**	**298.8**	**0.0036**	**CT(FAD–PAL)**
5	4.36	284.4	0.0004	LE(FAD)
6	4.50	275.6	0.0012	CT(FAD–PAL)
7	4.82	257.3	0.0022	LE(FAD)
8	4.90	253.1	0.1006	LE(FAD)
9	5.02	247.0	0.0379	Mixed LE/CT state
10	5.13	241.7	0.0327	Mixed LE/CT state

Benchmark calculations

To choose the DFT functional for absorption spectrum calculations of FAD in CvFAP a benchmark calculation was performed in a QM/MM setup, including in the QM region only the lumiflavin moiety of FAD. The lumiflavin structure was optimized in QM/MM environment for each functional and then TD-DFT calculations were performed using the optimized structure. This benchmark was done in a QM/MM framework, with the electrostatic embedding scheme, hence a direct comparison with experimental data is possible. Here the mean absolute error (MAE) on peak position and the relative intensities between the experimental and QM/MM spectrum were reported. Among the tested functionals, B3LYP and M06 exhibit the best overall agreement with the experimental data [117], yielding the lowest mean average error (MAE) for peak positions and lowest error on relative intensities. All calculations were performed with def2-TZVP basis set.

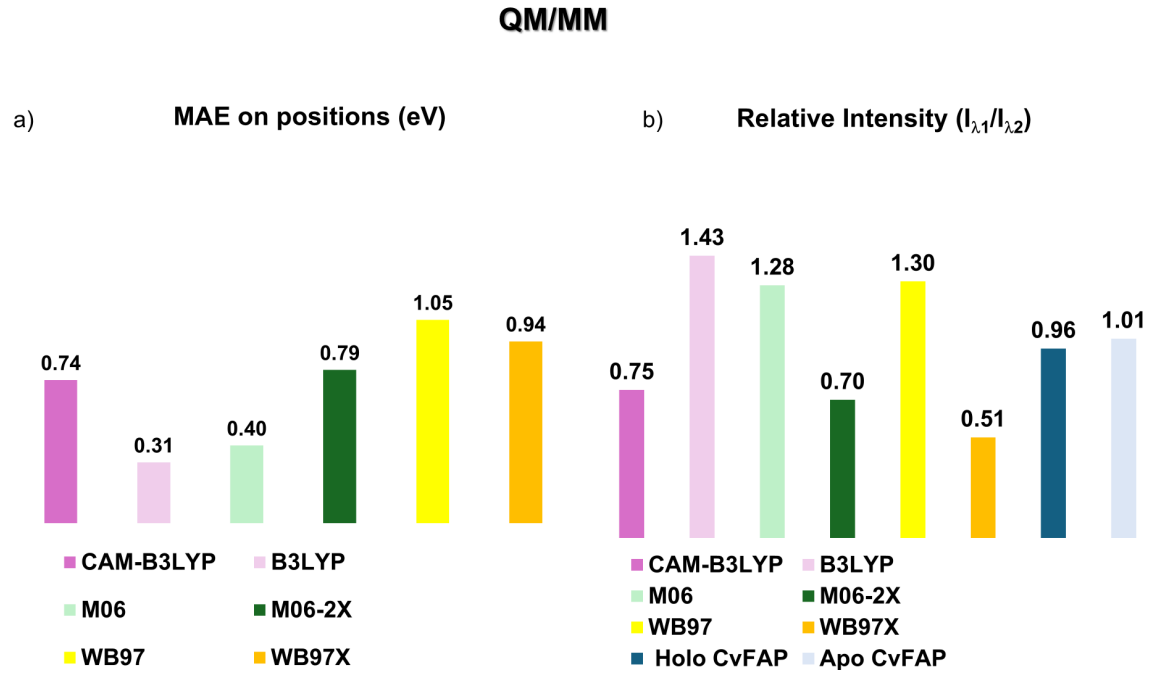


Figure S1: (a) Mean absolute error (MAE) of the absorption band positions with respect to the experimental values, comparing different DFT functionals within a QM/MM framework (QM region: lumiflavin). (b) Comparison between the experimental and computed relative intensities of the absorption bands. All calculations were performed with def2-TZVP basis set.

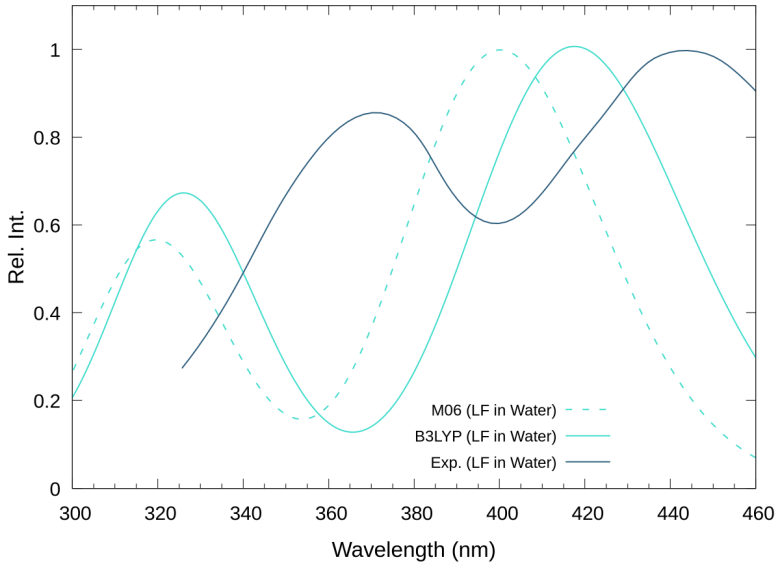


Figure S2: Comparison between the experimental absorption spectrum of lumiflavin in water (blue line) with the spectra obtained through the MD-PMM using two different DFT functionals: M06 (dashed cyan lines) and B3LYP (cyan lines). B3LYP achieves a closer match to the experimental results, both in peak positions and relative intensities.

Unperturbed Calculations

Table S3: Gas-phase TD-DFT transition energies, transition dipole moments and dipole moments for the lumiflavin planar structure

State	D_X	D_Y	D_Z
0	−3.4095	−0.8903	−1.1908
1	−3.6091	−1.4871	−1.0245
2	−1.3661	−0.3504	−0.4813
3	−2.1677	−1.1352	−0.5118
4	−4.9705	−1.0095	−1.8601
5	−0.6435	−0.8873	0.0850
6	−0.6244	−0.2409	−0.1866
7	−2.8107	−1.2769	−0.7468
8	−4.1840	−1.2863	−1.3774
9	−3.2301	−1.0794	−1.0259
10	−2.6266	−1.0901	−0.7428

Trans.	E (eV)	λ (nm)	f_{osc}	D_X	D_Y	D_Z
$0 \rightarrow 1$	3.0207	410.4	0.1987	−1.5014	−0.4021	−0.5193
$0 \rightarrow 2$	3.1549	393.0	0.0003	0.0244	−0.0246	−0.0561
$0 \rightarrow 3$	3.2896	376.9	0.0003	−0.0274	0.0192	0.0449
$0 \rightarrow 4$	3.8356	323.2	0.1447	1.0975	−0.1330	0.5637
$0 \rightarrow 5$	3.9196	316.3	0.0000	0.0019	−0.0020	−0.0038
$0 \rightarrow 6$	4.0660	304.9	0.0103	0.0705	0.2991	−0.0972
$0 \rightarrow 7$	4.5964	269.7	0.0002	−0.0158	0.0151	0.0332
$0 \rightarrow 8$	4.7042	263.6	0.0834	−0.7582	0.0690	−0.3796
$0 \rightarrow 9$	4.8707	254.6	0.5862	1.7597	−0.7333	1.1305
$0 \rightarrow 10$	5.0126	247.3	0.0232	0.3956	0.0032	0.1805

Table S4: Gas-phase TD-DFT transition energies, transition dipole moments and dipole moments for the partially bent (10^o bending) lumiflavin structure.

State	D_X	D_Y	D_Z
0	−2.2910	−0.6134	−2.8341
1	−2.6157	−1.1362	−2.7210
2	−0.8975	−0.2772	−1.2160
3	−1.7445	−0.8742	−1.5849
4	−3.0600	−0.7015	−4.1527
5	−0.7722	−0.7572	−0.4830
6	−0.3558	−0.3397	−0.5273
7	−2.2766	−0.9747	−2.3378
8	−2.7128	−0.9378	−3.1447
9	−2.3108	−0.7888	−2.6073
20	−1.9152	−0.8415	−1.9433

Trans.	E(eV)	λ(nm)	f_{osc}	D_X	D_Y	D_Z
$0 \rightarrow 1$	3.0151	411.2	0.1856	0.9979	0.2681	1.2020
$0 \rightarrow 2$	3.1527	393.3	0.0011	−0.1125	0.0139	−0.0372
$0 \rightarrow 3$	3.3038	375.3	0.0103	−0.1736	−0.0766	−0.3020
$0 \rightarrow 4$	3.8292	323.8	0.1478	0.5106	−0.1518	1.1365
$0 \rightarrow 5$	3.9091	317.2	0.0032	0.1055	0.0698	0.1302
$0 \rightarrow 6$	4.0787	304.0	0.0096	−0.1416	−0.2665	0.0672
$0 \rightarrow 7$	4.5792	270.8	0.0045	0.1214	0.0326	0.1572
$0 \rightarrow 8$	4.7239	262.5	0.0923	0.3138	−0.0763	0.8328
$0 \rightarrow 9$	4.8883	253.6	0.5534	−0.5820	0.6978	−1.9481
$0 \rightarrow 10$	4.9949	248.2	0.0105	−0.1912	0.0174	−0.2216

Table S5: Gas-phase TD-DFT transition energies, transition dipole moments and dipole moments for the bent (20^o bending) lumiflavin structure.

State	D_X	D_Y	D_Z
0	−3.4089	−0.7895	−0.9205
1	−3.5340	−1.3847	−0.8667
2	−1.4718	−0.3462	−0.5239
3	−2.1652	−1.0721	−0.4234
4	−5.0105	−0.8857	−1.6017
5	−0.6043	−0.9357	−0.1799
6	−0.5323	−0.3197	−0.3856
7	−3.2333	−1.2399	−0.7721
8	−3.7007	−1.1576	−0.9756
9	−3.3026	−0.9936	−0.8231
10	−2.3657	−1.0684	−0.5006

State	E(eV)	λ(nm)	f_{osc}	D_X	D_Y	D_Z
$0 \rightarrow 1$	2.9890	414.8	0.1655	1.4229	0.3409	0.3460
$0 \rightarrow 2$	3.1397	394.9	0.0006	−0.0585	−0.0100	0.0649
$0 \rightarrow 3$	3.3037	375.3	0.0225	0.4831	0.1047	0.1825
$0 \rightarrow 4$	3.7942	326.8	0.1548	−1.1662	0.1487	−0.5325
$0 \rightarrow 5$	3.9115	317.0	0.0005	−0.0529	−0.0383	−0.0266
$0 \rightarrow 6$	4.0566	305.6	0.0111	−0.0130	0.3289	−0.0592
$0 \rightarrow 7$	4.5348	273.4	0.0082	−0.2358	−0.0642	−0.1179
$0 \rightarrow 8$	4.7249	262.4	0.0835	−0.7211	0.1180	−0.4325
$0 \rightarrow 9$	4.8933	253.4	0.5058	1.7155	−0.7668	0.8295
$0 \rightarrow 10$	4.9162	252.2	0.0062	−0.0993	0.0913	−0.1829

Dihedral distribution functions

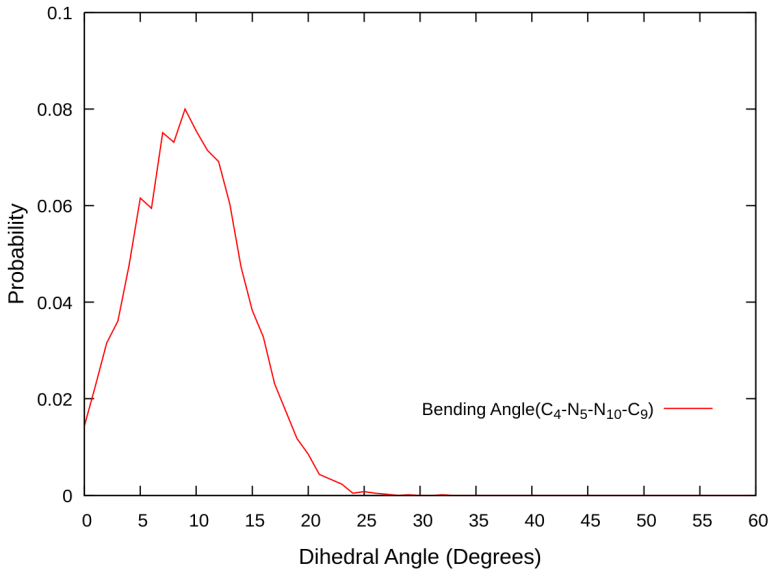


Figure S3: Distribution of the lumiflavin dihedral angle (C_4–N_5–N_{10}–C_9) sampled along the Molecular Dynamics (MD) trajectory of CvFAP in the *apo* state. The distribution displays a main peak centered at approximately 9°.

Unscaled MD-PMM calculations

Table S6: Unscaled MD-PMM calculations.

Conformation	λ_1 (nm)	λ_2 (nm)
Bent (20°)	434	334
Part. Bent	431	331
Planar	431	331
Weighted	432	332

MD-PMM calculations with different topology

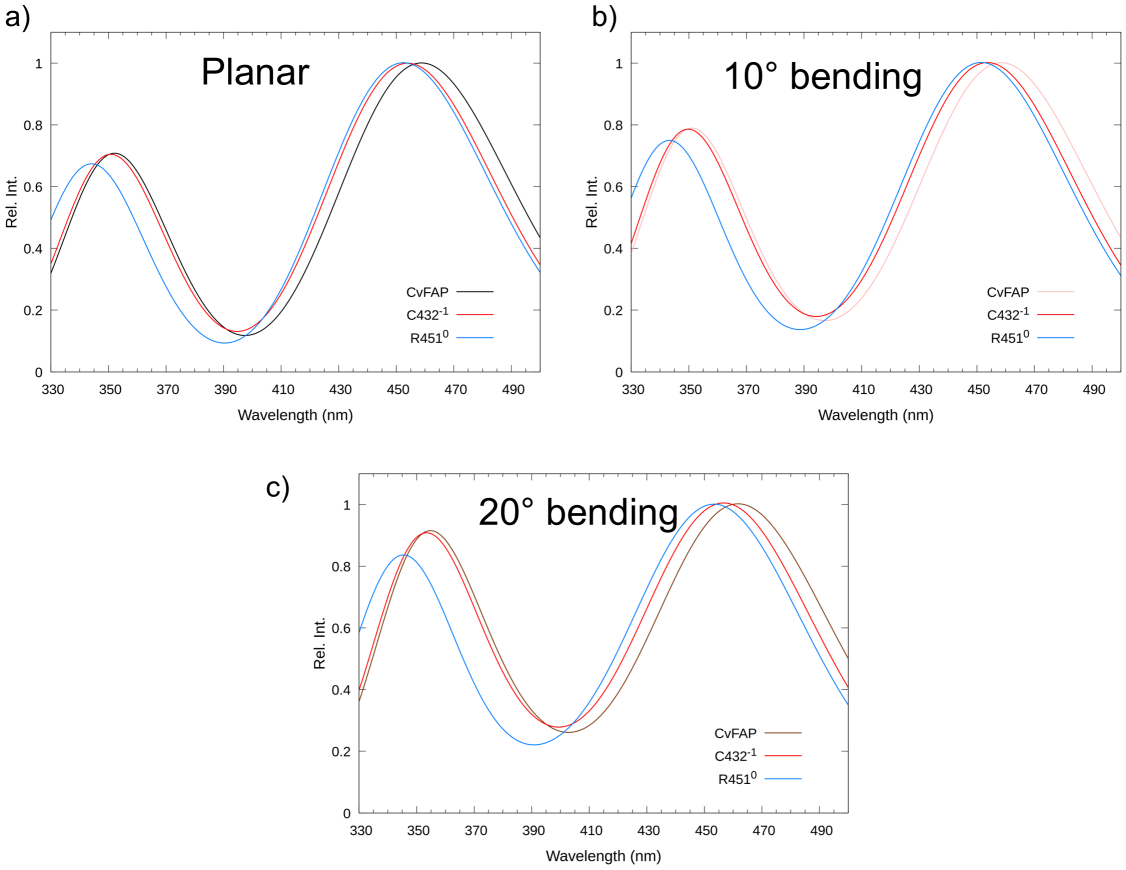


Figure S4: MD-PMM Absorption spectra of FAD in CvFAP computed employing either the standard charges or a modified topology, in which Arg451 was set to the neutral state (0) and Cys432 to the anionic state (−1). The calculations were carried out for three distinct lumiflavin conformations: (a) planar, (b) partially bent (10°), and (c) bent (20°.)

5.3 List of publications

[1] Matteo Farina, Matteo Capone, Enrico Bodo, Richard H. Fish, Massiliano Aschi, Alessandro Marrone, and Isabella Daidone. Mechanisms in the Synthesis of *(S)*-Alcohols with 1,4-NADH Biomimetic Co-factor N-Benzyl-1,4-dihydronicotinamide using Horse Liver Alcohol Dehydrogenase: A Hybrid Computational Study. *ChemBioChem*, 26(3), 2025.

[2] Matteo Farina, Flaminia Rondino, Andrea Lapi, Mauro Falconieri, Serena Gagliardi, Isabella Daidone, Caterina Fraschetti, Enrico Bodo, and Antonello Filippi. Applying computational spectroscopy methods to raman spectra of dicationic, imidazolium-based, ionic liquids. *J. of Phys.Chem. B*, 128(43):10650–10660, 2024.

[3] Alice Vetrano, Matteo Capone, Matteo Farina, Francesco Gabriele, Nicoletta Spreti, and Isabella Daidone. Unveiling cofactor inhibition mechanisms in horse liver alcohol dehydrogenase: An allosteric driven regulation. *Bioorg. Chem.*, 153:107932, 2024.

[4] Gianluca dell'Orletta, Nico Di Fonte, Matteo Farina, and Isabella Daidone. Insights into Substrate Protonation and Solvent Accessibility in the Active Site of Fatty Acid Photodecarboxylase. *J. Phys. Chem. Lett.*, 16, 49, 12538–12544, 2025

[5] Matteo Farina, Gianluca dell'Orletta, Enrico Bodo and Isabella Daidone. On the origin of the red-shifted flavin absorption spectra in fatty acid photodecarboxylase. Accepted to J. Phys. Chem. B.

5.4 List of abbreviations

AMBER Assisted Model Building with Energy Refinement

APO Apoprotein

BET Back electron transfer

BNAH Benzyldihydronicotinamide

CHARMM Chemistry at Harvard Macromolecular Mechanics

CT Charge-Transfer

CVFAP Chlorella variabilis Fatty Acid Photodecarboxylase

DDF Dihedral Distribution Function

DFT Density Functional Theory

FAP Fatty Acid Photodecarboxylase

GAFF General Amber Force Field

GFN-FF Geometries, Frequencies, and Non-covalent interactions Force Field

HLADH Horse Liver Alcohol Dehydrogenase

HOLO Holoprotein

HT Hydride transfer

ILS Ionic Liquids

IR Infrared

MD Molecular Dynamics

MD-PMM Molecular Dynamics Perturbed Matrix Method

MM Molecular Mechanics

1,4-NADH Nicotinamide Adenine Dinucleotide

NTC Neutral Triple Cluster

NTOs Natural Transition Orbitals

ONIOM Our Own N-layered Integrated Molecular Orbital and Molecular Mechanics

PCET Proton-Coupled Electron Transfer

PES Potential Energy Surface

PT Proton Transfer

TD-DFT Time-Dependent Density Functional Theory

UV-Vis Ultraviolet–Visible

xTB (GFN2-xTB) Semiempirical Extended Tight Binding

5.5 Acknowledgments

A special thanks to Professor E. Bodo, who has brilliantly guided me in achieving all of my academic goals. I also wish to thank Professor I. Daidone and Professor M. Aschi, who have supported and assisted me with care and kindness throughout my doctoral journey. Thank you also to all my colleagues of the groups of Prof. Bodo : Stefano, Adriano, Eleonora, Federica, Chiara, Francesca, Vanessa, Alessandro and Matteo Capone who supported me during the whole Ph.D.

Additionally, I want to show my gratitude to the most important persons of my life: my family, my friends, and all the important people in my life who have stood by me during this Ph.D. journey in particular Lorenzo, Cinzia and Pierluigi.

And finally the most special thank you goes to Özge, my fiancée and future wife, who has made my life better by being on my side in both difficult and happy moments.

We acknowledge ISCRA for awarding this project access to the LEONARDO supercomputer, owned by the EuroHPC Joint Undertaking, hosted by CINECA (Italy)[151].

We acknowledge the use of ChatGPT (https://chatgpt.com/) to identify improvements in the writing style

List of Figures

1.1 Key features of complex systems: large size, significant interactions between the components, structural complexity, and conformational changes. All these features result in a non-linear behaviour. 7

1.2 A summary of possible applications of biocatalysis. 9

1.3 Redox states of FAD (a) and NAD^+ (b). 10

1.4 Structure of 1-benzyl-1,4-dihydronicotinamide (BNAH). In the work of Fish et al. [14] this molecule was used as biomimetic cofactor replacing the natural 1,4-NADH for the HLADH-catalyzed reduction of ketones to *(S)*-alcohol. 12

1.5 Structures of the most common flavins: a) lumiflavin, b) riboflavin c) FMN and d) FAD. Flavins are characterized by a common isoalloxazine ring, differing in the various functional substituents linked to that core. 14

1.6 Basic QM/MM partitioning. The most chemically relevant part of the system is included (e.g. the lumiflavin moiety of FAD cofactor) in the QM part, while the MM part encompasses the remaining of the system (usually protein and solvent). 18

2.1 ONIOM Three Layers system division The QM1 is the region where the process of interest occurs, generally modelled with high level DFT. The QM2 region is the "surrounding" of the QM1 region, generally treated at semiempirical level, while the MM region is the external environment. 27

2.2 ONIOM Microiteration algorithm 29

2.3 Workflow of a typical MD-PMM calculation The key step is the diagonalization of the perturbed Hamiltonian matrix, including the effect of the environmental perturbation at each MD frame. In green we have highlighted the QM/MM step introduced in this work. 33
2.4 Atom numbers assigned to the lumiflavin moiety of the FAD cofactor . 39

3.1 Reaction mechanism of 2-pentanone reduction catalyzed by HLADH in the presence of the native 1,4-NADH cofactor [111, 17, 15]. In the first step (i), a direct hydride transfer occurs from the C_4 atom of NADH to the C_2 atom of the substrate, forming the corresponding *(S)*-pentan-2-olate intermediate. In the following step (ii), His51 is protonated by the bulk. Subsequently (iii), a proton is transferred from $His51^+$ to the *(S)*-pentan-2-olate, yielding the corresponding *(S)*-alcohol. This step is mediated by a hydrogen-bond network connecting $His51^+$ to the *(S)*-pentan-2-olate through the riboside moiety of NADH. 42
3.2 Molecular structures of BNAH (1-benzyl-1,4-dihydronicotinamide) in its reduced (1) and oxidized (2^+) forms. 43
3.3 Alternative mechanism proposed by A. Marrone et al. in the presence of BNAH biomimetic [15, 16]. The proposed mechanism involves an initial dissociation of His67 from the Zn(II) center (i) yielding the η^2-5,6-1,4-dihydronicotinamide–Zn(II) complex (η^2-complex). The formation of this intermediate is followed by a direct hydride transfer from C_4 of BNAH to C_2 of the substrate (ii). The catalytic cycle proceeds with the protonation of His51 by the bulk solvent (iii), which allows the final proton transfer step (iv) required to form the (*S*)-alcohol. 45

3.4 a) Active site of HLADH in the HIE51 configuration. The catalytic Zn^{2+} ion is coordinated by Cys174, Cys46, His67, and the oxygen atom of 2-pentanone. The 2-pentanone substrate is positioned in a pro-*S* orientation, while the natural 1,4-NADH cofactor is replaced by BNAH (denoted as 1). Here the hydrogen bond between Ser48 and the carbonyl oxygen of 2-pentanone is shown with red dotted line, the C_4-C_2 distance is highlighted in blue and the $C_4 - H_4 - C_2$ angle is denoted as α. The coordination of the catalytic Zn^{2+} is depicted with thick black lines. b) RMSD (Root Mean Square Deviation) calculated on the protein backbone. c) Time evolution of the C_4(BNAH)-C_2(2-pentanone) distance. d) Time evolution of the C_4(BNAH)-H_4-C_2(2-pentanone) angle. 48

3.5 QM/MM optimized structures corresponding to the initial points used in the ONIOM scans are shown. The QM region in the ONIOM scans, corresponding to the entire HLADH active site, is depicted. Here all amino acids are saturated at the Cα atom (Ser48,His51,Cys174,Cys46) while the substrate (2-pentanone) and the cofactor (BNAH) are fully included in the QM Part. Three different configurations were tested: (a) Neutral His51 (HIE51); (b) Cationic His51 (HIP51); (c) Cationic His51 (HIP51) with an additional water molecule inserted between His51 and Ser48. Hydrogen bonds between Ser48 and 2-pentanone are highlighted in red, while those between water and other residues are highlighted in blue. The H_4(BNAH)-C_2(2-pent.) distance (reaction coordinate) for each structure is indicated with green lines: (a) 2.63 Å, (b) 2.59 Å, and (c) 2.52 Å. Structure reported in panel (a) correspond to the QC region in the subsequent MD-PMM calculation. 50

3.6 Unperturbed energy barrier along each mechanistic path calculated at B3LYP/631G* level of theory. The blue lines in the mechanisms V and VI indicate the value of the energy barrier for the histidine detachment step. 53

3.7 MD-PMM reaction pathways, comparing the Gibbs' free energy barriers of the reaction model I (a) and reaction model V (b). R represents the reagent structure, H.E.P. is the highest energy point, and P is the product structure. In the mechanism (b) also the geometry corresponding to the η^2-complex formed after His67 detachment is represented. Crucial bond distances, related to the reaction coordinates, for both reaction pathways are highlighted in red. In the mechanism (b) the His67 contribution to the energy barrier is highlighted by a green curly bracket. 56

3.8 a) Active site of the HIP51 configuration. Here the hydrogen bond between Ser48 and the carbonyl oxygen of 2-pentanone is shown with red dotted line, the C_4-C_2 distance is highlighted in blue and the $C_4{-}H_4{-}C_2$ angle is denoted as α. The coordination of the catalytic Zn(II) is depicted with thick black lines. b) RMSD (Root Mean Square Deviation) calculated on the protein backbone. c) Time evolution of the C_4(BNAH)-C_2(2-pentanone) distance. d) Time evolution of the C_4(BNAH)-H_4-C_2(2-pentanone) angle. 58

3.9 a) Reaction center of HLADH with charged His51 (HIP51). This snapshot from MD clearly shows a water molecule (Wat.) bridging His51 and Ser48, whose interactions are highlighted by blue lines. b) The hydrogen bond network involving two water molecules. c) Scheme of the ONIOM scan for the simulation of the proton-transfer chain. The reaction coordinate H(Ser48)–O(*(S)*-2-pentan-2-olate) is indicated by a red arrow, while the dotted red lines highlights the hydrogen bond network. 59

3.10 Proposed mechanism for the HLADH-catalyzed reduction of 2-pentanone with the biomimetic co-factor, BNAH. In the first step (i), a direct hydride transfer occurs from the C_4 atom of BNAH to the C_2 atom of the substrate, forming the corresponding *(S)*-pentan-2-olate intermediate. In the following step (ii), His51 is protonated by the bulk. Subsequently (iii), a proton is transferred from $His51^+$ to the *(S)*-pentan-2-olate, yielding the corresponding *(S)*-alcohol. This last step is mediated by a a water molecule connecting $His51^+$ to Ser48 through an hydrogen bond network. 61

3.11 Experimental spectra of FAD in *apo* (purple) [118] and *holo* (green) [116] CvFAP compared to the experimental spectrum of lumiflavin in water (blue) [121]. The two FAD main peaks in CvFAP, at 467 and 390 nm, are red-shifted by 21-23 nm compared to the peaks of lumiflavin in water (443 and 370 nm). The cyan box highlight the spectral range (450–460 nm) where generally the first FAD absorption peak is observed in other flavoproteins. 63

3.12 QM/MM-optimized structures of the CvFAP active site for the two substrate protonation states: QM region (a) and active site (c) of the deprotonated palmitic acid (palmitate) configuration; (b) QM region and active site (d) of the neutral palmitic acid configuration. The QM region includes: the head of the palmitic/palmitate moiety, the lumiflavin, two catalytic water molecules (Wat1 and Wat2) and the side chain of R451. The active site depiction highlights also the MM residues Q486 (side chain), C432 (side chain), S573, S574, and N575 that establish direct h-bond with QM atoms. 65

3.13 NTOs of respectively locally excited (LE) (a) and (b) (CT) state for the configuration containing the palmitate substrate. Calculations were performed at the CAM-B3LYP/def2-TZVP level of theory. Here only QM atoms, in the ground-state geometry, are depicted. 66

3.14 QM/MM optimized structure of the first CT state and LE state at the CAM-B3LYP/def2-TZVP level of theory. QM region (a) and active site (c) of the optimized CT state geometry; QM region (b) and active site (d) of the optimized LE state geometry. 67

3.15 Correlation between the molecular conformation of lumiflavin and its absorption spectrum in gas-phase. a) Optimized conformations of lumiflavin, at the B3LYP/def2-TZVP level, obtained at different values of the C_4–N_5–N_{10}-C_9 dihedral angle (highlighted with red dashed lines): bent (20°), partially bent (10°), and planar (0°) structure. b) TD-DFT absorption peaks (nm) and normalized absorption intensities for the first two bright electronic states of the three conformers calculated in gas-phase at the same level of theory. A scaling factor of 0.94 was applied to the energies (in eV). 72

3.16 MD-PMM absorption spectra of the three conformations of FAD in CvFAP: planar (black), partially bent (pink), and bent (brown). 74

3.17 Comparison of MD-PMM (a) and experimental spectra (b) of lumiflavin in water and in CvFAP. The MD-weighted PMM spectrum in CvFAP (purple) closely matches the experimental one, reproducing both the relative peak intensity (0.81 vs. 1.01) and the red shift of the first absorption maximum (16 nm vs. 23 nm) compared to water (blue). . . 75

3.18 Comparison between the QM/MM and the MD-PMM spectra of FAD calculated in CvFAP. In both calculations only lumiflavin was included in the QM part. A scaling factor of 0.94 was applied to energies. 76

3.19 Comparison between the MD-weighted spectrum in CvFAP obtained using MD-PMM calculations with the standard charges and those with altered charges, namely neutral Arg451 and negatively charged (-1) Cys432. In both cases, the charge alterations result in a blue shift, rather than in a red shift, more pronounced for the neutral Arg451. . . 79

3.20 QM/MM-optimized structures for the reagent and product structures: (a) QM region of the reagent structure; (b) QM region of the product configuration; (c) reagent active site; (d) product active site. The QM region, in the reagent includes the head of the palmitate ion (-CH_2-CH_2-CO_2^- fragment), the lumiflavin moiety of FAD, two catalytic water molecules (Wat1 and Wat2) and the side chain of C432 and R451. In the product region, the bicarbonate ion and the alkane head RH (-CH_2-CH_3 fragment) replace the head of the palmitate ion and the Wat1. The active site representation includes also the MM residues Q486 (side chain) S573, S574, and N575. 80

3.21 (a) Comparison of the FAD absorption peaks between the QM/MM spectra of the product (dashed line) and the reactant structure (solid line). (b) Experimental spectra of CvFAP [117] in the dark state (solid line) and after illumination (dashed line). Calculations were performed in a QM/MM setup at the B3LYP def2-TZVP level of theory. Atoms included in the QM are reported in figure 3.20. A scaling factor of 0.94 was applied on energies. 81

3.22 Characterization of the molecular orbitals involved in the first bright transition: a) Comparison of the first bright transition energy gap in reagent (red orbitals) and product structures (blue orbitals). The product shows a smaller gap due to destabilization of the initially occupied orbital during formation. Corresponding molecular orbitals are displayed (blue = negative, red = positive). b) NTOs related to the first bright transition in both structures. 83

4.1 Summary of the computational workflow employed in this thesis. 89

5.1 General structure of the dicationic ionic liquids studied in the work. Here n represent the number of C atoms in the alkyl chain linking the two 3-methylimidaziolium rings. 91

5.2 a) Structure of different intramolecular conformers for both cation and anion. b) Dihedral distribution functions for the three cations: C2, C3 and C4. c) Dihedral distribution function for the anion $TFSI^-$. The dihedral distribution function of the anion does not show any significant variations in all four systems. 94

5.3 Comparison between simulated and experimental spectra of $[C3(mim)_2](TFSI)_2$ in the low frequency zones (260-440cm^{-1}). Peaks related to different conformers are highlighted. The applied scaling factor is 1.04 for the anion and 0.97 for the cation. 96

5.4 Comparison between simulated and experimental spectra of (a) $[C1(mim)_2](TFSI)_2$, (b) $[C2(mim)_2](TFSI)_2$, (c) $[C3(mim)_2](TFSI)_2$ and (d) $[C4(mim)_2](TFSI)_2$ in the fingerprint zone (400-1800cm^{-1}). Peaks related to different conformers are highlighted. The applied scaling factor is 1.04 for the anion and 0.97 for the cation. 97

5.5 Comparison between simulated and experimental spectra of: (a) $[C1(mim)_2](TFSI)_2$, (b) $[C2(mim)_2](TFSI)_2$, (c) $[C3(mim)_2](TFSI)_2$ and (d) $[C4(mim)_2](TFSI)_2$ in the high frequency zone (2800-3200cm^{-1}). The applied scaling factor is 0.96 for the cation. No peaks related to $TFSI^-$ are present in that zone. 99

S1 (a) Mean absolute error (MAE) of the absorption band positions with respect to the experimental values, comparing different DFT functionals within a QM/MM framework (QM region: lumiflavin). (b) Comparison between the experimental and computed relative intensities of the absorption bands. All calculations were performed with def2-TZVP basis set. 104

S2 Comparison between the experimental absorption spectrum of lumiflavin in water (blue line) with the spectra obtained through the MD-PMM using two different DFT functionals: M06 (dashed cyan lines) and B3LYP (cyan lines). B3LYP achieves a closer match to the experimental results, both in peak positions and relative intensities. 105

S3 Distribution of the lumiflavin dihedral angle (C_4–N_5–N_{10}–C_9) sampled along the Molecular Dynamics (MD) trajectory of CvFAP in the *apo* state. The distribution displays a main peak centered at approximately 9°. 108

S4 MD-PMM Absorption spectra of FAD in CvFAP computed employing either the standard charges or a modified topology, in which Arg451 was set to the neutral state (0) and Cys432 to the anionic state (−1). The calculations were carried out for three distinct lumiflavin conformations: (a) planar, (b) partially bent (10°), and (c) bent (20°.) 109

List of Tables

3.1 HLADH simulated PES scans. Models I and V starts from structure shown in Figure 3.5 (a); starting structure for model II is shown in panel (b) and the starting point for models III, IV, and VI is shown in panel (c). 51

3.2 a) Positions of the main UV-Vis absorption peaks of FAD within *apo* CvFAP [118] and *holo* CvFAP [117] and of lumiflavin in water [121]; and (b) their relative intensity. 63

3.3 a) TD-DFT gas-phase positions of the main lumiflavin UV-Vis absorption peaks. and b) their relative intensity, for lumiflavin in various conformations at the B3LYP/def2-TZVP level of theory. Energies were corrected by a scaling factor of 0.94. Data were compared with experimental spectrum in CvFAP [118] 71

3.4 MD-PMM positions (nm) of the main lumiflavin UV-Vis absorption peaks and their relative intensity (I_{rel}) for the three FAD conformations. 74

3.5 Positions of the UV-Vis absorption peaks and relative intensities (I_{rel}) for FAD in CvFAP and lumiflavin in water, based on MD-PMM calculations and experimental data. 75

5.1 Assignment of the vibrational modes of the anion TFSI. 98

5.2 Assignment for the main vibrational modes of cations Cn (n=1-4) in the fingerprint zone. 98

5.1 Detailed assignment of the normal vibrational modes of cations Cn (n=1-4), for the high frequency zone. 100

S1 Energy, oscillator strength (f.osc), wavelength, and nature of the first ten excited states of the QM region calculated at the CAM-B3LYP/def2-TZVP level of theory for the RCOOH configuration. In bold it is reported also the CT (PAL–FAD) that is the 19^{th} state. 103

S2 Energy, oscillator strength (f.osc), wavelength, and nature of the first ten excited states of the QM region calculated at the CAM-B3LYP/def2-TZVP level of theory for the $RCOO^-$ configuration. In bold it is reported the CT (PAL–FAD) that is the 4^{th} state. 103

S3 Gas-phase TD-DFT transition energies, transition dipole moments and dipole moments for the lumiflavin planar structure 105

S4 Gas-phase TD-DFT transition energies, transition dipole moments and dipole moments for the partially bent (10^o bending) lumiflavin structure. 106

S5 Gas-phase TD-DFT transition energies, transition dipole moments and dipole moments for the bent (20^o bending) lumiflavin structure. 107

S6 Unscaled MD-PMM calculations. 108

Bibliography

[1] Avi Ma'ayan. Complex systems biology. *J. R. Soc. Interface*, 14(134):20170391, 2017.

[2] Jordan Rozum and Réka Albert. Leveraging network structure in nonlinear control. *npj Syst. Biol. Appl.*, 8(1):36, 2022.

[3] Dong Yi, Thomas Bayer, Christoffel P. S. Badenhorst, Shuke Wu, Mark Doerr, Matthias Höhne, and Uwe T. Bornscheuer. Recent trends in biocatalysis. *Chem. Soc. Rev.*, 50(14):8003–8049, 2021.

[4] Adam O'Connell, Amber Barry, Ashleigh J. Burke, Amy E. Hutton, Elizabeth L. Bell, Anthony P. Green, and Elaine O'Reilly. Biocatalysis: landmark discoveries and applications in chemical synthesis. *Chem. Soc. Rev.*, 53(6):2828–2850, 2024.

[5] Joel M. Hawkins and Timothy J. N. Watson. Asymmetric Catalysis in the Pharmaceutical Industry. *Angew. Chemie Int. Ed.*, 43(25):3224–3228, 2004.

[6] Francesco Falcioni, Luke Humphreys, Richard C. Lloyd, Hao Wu, Isamir Martinez, Jonathan Jones, Shane McKenna, Katharina Neufeld, Ryan M. Phelan, Tay Rosenthal, Christophe J. Szczepaniak, Kumiko Yamamoto, Scott P. France, and Anna Fryszkowska. The Evolving Landscape of Industrial Biocatalysis in Perspective from the ACS Green Chemistry Institute Pharmaceutical Roundtable. *ACS Catal.*, pages 10780–10794, 2025.

[7] Sara Amjad, Sabah Nisar, Ajaz A. Bhat, Ab Rauf Shah, Michael P. Frenneaux, Khalid Fakhro, Mohammad Haris, Ravinder Reddy, Zoltan Patay, Joseph Baur,

and Puneet Bagga. Role of NAD+ in regulating cellular and metabolic signaling pathways. *Mol. Metab.*, 49:101195, 2021.

[8] Teresa Anna Giancaspero, Vittoria Locato, and Maria Barile. A Regulatory Role of NAD Redox Status on Flavin Cofactor Homeostasis in cervisiae Mitochondria. *Oxid. Med. Cell. Longev.*, 2013:1–16, 2013.

[9] Alexey Aleksandrov. A Molecular Mechanics Model for Flavins. *J. Comput. Chem.*, 40(32):2834–2842, 2019.

[10] Rosa María Martínez-Espinosa, Micaela Giani Alonso, Lorena Simó Cabrera, Guillermo Martínez, and Carmen Pire Galiana. Industrial applications of enzymes from haloarchaea. In *Extrem. Their Ind. Appl.*, pages 289–320. Elsevier, 2022.

[11] Alice Vetrano, Matteo Capone, Matteo Farina, Francesco Gabriele, Nicoletta Spreti, and Isabella Daidone. Unveiling cofactor inhibition mechanisms in horse liver alcohol dehydrogenase: An allosteric driven regulation. *Bioorg. Chem.*, 153:107932, 2024.

[12] Musa M. Musa and Robert S. Phillips. Recent advances in alcohol dehydrogenase-catalyzed asymmetric production of hydrophobic alcohols. *Catal. Sci. Technol.*, 1:1311–1323, 2011.

[13] Norman J. Oppenheimer and Gary T. M. Henehan. Horse Liver Alcohol Dehydrogenase-Catalyzed Aldehyde Oxidation. pages 407–415. 1995.

[14] Prof. R. Fish and H. C. Lo. Biomimetic NAD+ Models for Tandem Cofactor Regeneration, Horse Liver Alcohol Dehydrogenase Recognition of 1,4-NADH Derivatives, and Chiral Synthesis. *Angew. Chem. Int. Ed.*, 41(3):478–481, 2002.

[15] Matteo Farina, Matteo Capone, Enrico Bodo, Richard H. Fish, Massiliano Aschi, Alessandro Marrone, and Isabella Daidone. Mechanisms in the Synthesis of (S)-Alcohols with 1,4-NADH Biomimetic Co-factor N-Benzyl-1,4-

dihydronicotinamide using Horse Liver Alcohol Dehydrogenase: A Hybrid Computational Study. *ChemBioChem*, 26(3), 2025.

[16] Alessandro Marrone and Richard H. Fish. DFT Mechanism Studies: Biomimetic 1,4-NADH Chemoselective, Co-factor Regeneration with [Cp*Rh(bpy)H]+, in Tandem with the Biocatalysis Pathways of a Core Model of the (HLADH)-Zn(II) Mediated Enzyme, in the Enantioselective Reduction of Achiral Ketones to . *J. Organomet. Chem.*, 943:121810, 2021.

[17] Pratul K. Agarwal, Simon P. Webb, and Sharon Hammes-Schiffer. Computational Studies of the Mechanism for Proton and Hydride Transfer in Liver Alcohol Dehydrogenase. *J. Am. Chem. Soc.*, 122(19):4803–4812, 2000.

[18] G. Schneider, H. Eklund, E. Cedergren-Zeppezauer, and M. Zeppezauer. Structure of the complex of active site metal-depleted horse liver alcohol dehydrogenase and NADH. *EMBO J.*, 2(5):685–689, 1983.

[19] Alessandro Marrone and Richard H. Fish. Bioorganometallic Chemistry at the Interface with Biocatalysis: Chemoselective Reduction of Biomimetic NAD^+ Cofactors with $[Cp^*Rh(bpy)H]^+$, Tandem Catalysis with 1,4-NADH-Dependent Enzymes, Chiral Synthesis, Organotin Metabolites. *Organometallics*, 42(4):288–306, 2023.

[20] B. Entsch and D.P. Ballou. Flavins. In *Encycl. Biol. Chem.*, pages 309–313. Elsevier, 2013.

[21] Rajiv Kumar Kar, Veniamin A. Borin, Yonghong Ding, Jörg Matysik, and Igor Schapiro. Spectroscopic Properties of Lumiflavin: A Quantum Chemical Study. *Photochem. Photobiol.*, 95(2):662–674, 2019.

[22] Rajiv K. Kar, Anne-Frances Miller, and Maria-Andrea Mroginski. Understanding flavin electronic structure and spectra. *WIREs Comput. Mol. Sci.*, 12(2), 2022.

[23] Stefan Weber and Erik Schleicher, editors. *Flavins and Flavoproteins*, volume 1146. Springer New York, New York, NY, 2014.

[24] Fauzy Nasher, Aidan J. Taylor, Abdi Elmi, Burhan Lehri, Umer Z. Ijaz, Dave Baker, Richard Goram, Steven Lynham, Dipali Singh, Richard Stabler, David J. Kelly, Ozan Gundogdu, and Brendan W. Wren. MdaB and NfrA, Two Novel Reductases Important in the Survival and Persistence of the Major Enteropathogen Campylobacter jejuni. *J. Bacteriol.*, 204(1), 2022.

[25] Steven O. Mansoorabadi, Christopher J. Thibodeaux, and Hung-wen Liu. The Diverse Roles of Flavin CoenzymesNature's Most Versatile Thespians. *J. Org. Chem.*, 72(17):6329–6342, 2007.

[26] Murat Kılıç and Bernd Ensing. First and Second One-Electron Reduction of Lumiflavin in Water—A First Principles Molecular Dynamics Study. *J. Chem. Theory Comput.*, 9(9):3889–3899, 2013.

[27] Martin Grininger, Kornelius Zeth, and Dieter Oesterhelt. Dodecins: A Family of Lumichrome Binding Proteins. *J. Mol. Biol.*, 357(3):842–857, 2006.

[28] James A. Watson, Patrick M. McTamney, Jennifer M. Adler, and Steven E. Rokita. Flavoprotein Iodotyrosine Deiodinase Functions without Cysteine Residues. *ChemBioChem*, 9(4):504–506, 2008.

[29] P. C. Anderson. Identification of a 14mer RNA that recognizes and binds flavin mononucleotide with high affinity. *Nucleic Acids Res.*, 33(22):6992–6999, 2005.

[30] Peter Deppisch, Charlotte Helfrich-Förster, and Pingkalai R. Senthilan. The Gain and Loss of Cryptochrome/Photolyase Family Members during Evolution. *Genes (Basel).*, 13(9):1613, 2022.

[31] LO Essen and T Klar. Light-driven dna repair by photolyases. *Cellular and Molecular Life Sciences CMLS*, 63(11):1266–1277, 2006.

[32] Laura Zanetti-Polzi, Massimiliano Aschi, Andrea Amadei, and Isabella Daidone. Alternative Electron-Transfer Channels Ensure Ultrafast Deactivation of Light-Induced Excited States in Riboflavin Binding Protein. *J. Phys. Chem. Lett.*, 8(14):3321–3327, 2017.

[33] Yuejie Ai, Chaofeng Zhao, Jinlu Xing, Yang Liu, Zhangxia Wang, Jiaren Jin, Shuhua Xia, Ganglong Cui, and Xiangke Wang. Excited-State Decay Pathways of Flavin Molecules in Five Redox Forms: The Role of Conical Intersections. *J. Phys. Chem. A*, 122(40):7954–7961, 2018.

[34] Jernej Stare. Oxidation of Flavin by Molecular Oxygen: Computational Insights into a Possible Radical Mechanism. *ACS Omega*, 9(22):23431–23441, 2024.

[35] John TM Kennis and Marie-Louise Groot. Ultrafast spectroscopy of biological photoreceptors. *Curr. Opin. Struct. Biol.*, 17(5):623–630, 2007.

[36] Bin Liu, Hongtao Liu, Dongping Zhong, and Chentao Lin. Searching for a photocycle of the cryptochrome photoreceptors. *Curr. Opin. Plant Biol.*, 13(5):578–586, 2010.

[37] Claus Nielsen, Morten S. Nørby, Jacob Kongsted, and Ilia A. Solov'yov. Absorption Spectra of FAD Embedded in Cryptochromes. *J. Phys. Chem. Lett.*, 9(13):3618–3623, 2018.

[38] Yi Xu, Martin T. Peschel, Miriam Jänchen, Richard Foja, Golo Storch, Erling Thyrhaug, Regina de Vivie-Riedle, and Jürgen Hauer. Determining Excited-State Absorption Properties of a Quinoid Flavin by Polarization-Resolved Transient Spectroscopy. *J. Phys. Chem. A*, 128(19):3830–3839, 2024.

[39] Tatiana Domratcheva, Elisabeth Hartmann, Ilme Schlichting, and Tilman Kottke. Evidence for Tautomerisation of Glutamine in BLUF Blue Light Receptors by Vibrational Spectroscopy and Computational Chemistry. *Sci. Rep.*, 6(1):22669, 2016.

[40] Laura Zanetti-Polzi, Massimiliano Aschi, Isabella Daidone, and Andrea Amadei. Theoretical modeling of the absorption spectrum of aqueous riboflavin. *Chem. Phys. Lett.*, 669:119–124, 2017.

[41] Padmabati Mondal, Karno Schwinn, and Miquel Huix-Rotllant. Impact of the redox state of flavin chromophores on the UV–vis spectra, redox and acidity constants and electron affinities. *J. Photochem. Photobiol. A Chem.*, 387:112164, 2020.

[42] Hans Martin Senn and Walter Thiel. QM/MM Methods for Biomolecular Systems. *Angew. Chemie Int. Ed.*, 48(7):1198–1229, 2009.

[43] Lung Wa Chung, W. M. C. Sameera, Romain Ramozzi, Alister J. Page, Miho Hatanaka, Galina P. Petrova, Travis V. Harris, Xin Li, Zhuofeng Ke, Fengyi Liu, Hai-Bei Li, Lina Ding, and Keiji Morokuma. The ONIOM Method and Its Applications. *Chem. Rev.*, 115(12):5678–5796, 2015.

[44] Peter Macheroux. UV-Visible Spectroscopy as a Tool to Study Flavoproteins. In *Flavoprotein Protoc.*, pages 1–8. Humana Press, New Jersey.

[45] Ya-Ting Kao, Chaitanya Saxena, Ting-Fang He, Lijun Guo, Lijuan Wang, Aziz Sancar, and Dongping Zhong. Ultrafast Dynamics of Flavins in Five Redox States. *J. Am. Chem. Soc.*, 130(39):13132–13139, 2008.

[46] Oliver Carrillo-Parramon, Sara Del Galdo, Massimiliano Aschi, Giordano Mancini, Andrea Amadei, and Vincenzo Barone. Flexible and Comprehensive Implementation of MD-PMM Approach in a General and Robust Code. *J. Chem. Theory Comput.*, 13(11):5506–5514, 2017.

[47] Laura Zanetti-Polzi, Sara Del Galdo, Isabella Daidone, Marco D'Abramo, Vincenzo Barone, Massimiliano Aschi, and Andrea Amadei. Extending the perturbed matrix method beyond the dipolar approximation: comparison of different levels of theory. *Phys. Chem. Chem. Phys.*, 20(37):24369–24378, 2018.

[48] Andrea Amadei, Maira D'Alessandro, Marco D'Abramo, and Massimiliano Aschi. Theoretical characterization of electronic states in interacting chemical systems. *J. Chem. Phys.*, 130(8), 2009.

[49] Maira D'Alessandro, Massimiliano Aschi, Claudia Mazzuca, Antonio Palleschi, and Andrea Amadei. Theoretical modeling of UV-Vis absorption and emission spectra in liquid state systems including vibrational and conformational effects: The vertical transition approximation. *J. Chem. Phys.*, 139(11), 2013.

[50] Solène Moulin, Bertrand Légeret, Stéphanie Blangy, Damien Sorigué, Adrien Burlacot, Pascaline Auroy, Yonghua Li-Beisson, Gilles Peltier, and Fred Beisson. Continuous photoproduction of hydrocarbon drop-in fuel by microbial cell factories. *Sci. Rep.*, 9(1):13713, 2019.

[51] Kaihao Mou, Yue Guo, Weihua Xu, Danyang Li, Zhiguo Wang, and Qi Wu. Stereodivergent protein engineering of fatty acid photodecarboxylase for light-driven kinetic resolution of sec-alcohol oxalates. *Angewandte Chemie International Edition*, 63(10):e202318374, 2024.

[52] L. Monticelli and D. P. Tieleman. *Biomolecular Simulations:Methods and Protocols*, volume 924. Humana Press, 2013.

[53] Mark E Tuckerman. Ab initio molecular dynamics: basic concepts, current trends and novel applications. *J. Phys. Condens. Matter*, 14(50):R1297–R1355, 2002.

[54] Jae Woo Park, Rachael Al-Saadon, Matthew K. MacLeod, Toru Shiozaki, and Bess Vlaisavljevich. Multireference Electron Correlation Methods: Journeys along Potential Energy Surfaces. *Chem. Rev.*, 120(13):5878–5909, 2020.

[55] Rodney J. Bartlett and Monika Musiał. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.*, 79(1):291–352, 2007.

[56] Axel D. Becke. Perspective: Fifty years of density-functional theory in chemical physics. *J. Chem. Phys.*, 140(18), 2014.

[57] Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.*, 15(3):1652–1671, 2019.

[58] Gerrit Groenhof. *Introduction to QM/MM Simulations*. 2013.

[59] Anders S. Christensen, Tomáš Kubař, Qiang Cui, and Marcus Elstner. Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications. *Chemical Reviews*, 116(9):5301–5337, 2016.

[60] Christoph Bannwarth, Eike Caldeweyher, Sebastian Ehlert, Andreas Hansen, Philipp Pracht, Jakob Seibert, Sebastian Spicher, and Stefan Grimme. Extended tight-binding quantum chemistry methods. *WIREs Comput. Mol. Sci.*, 11(2), 2021.

[61] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinforma.*, 65(3):712–725, 2006.

[62] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25(9):1157–1174, 2004.

[63] Jing Huang and Alexander D. MacKerell Jr. Charmm36 all-atom additive protein force field: Validation based on comparison to nmr data. *Journal of Computational Chemistry*, 34(25):2135–2145, 2013.

[64] William L. Jorgensen and Julian Tirado-Rives. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988.

[65] Xiaoli Lu, Jinfeng Chen, and Jing Huang. The continuous evolution of biomolecular force fields. *Structure*, 2025.

[66] VS Sandeep Inakollu, Daan P Geerke, Christopher N Rowley, and Haibo Yu. Polarisable force fields: what do they add in biomolecular simulations? *Current Opinion in Structural Biology*, 61:182–190, 2020. Theory and Simulation Macromolecular Assemblies.

[67] Yue Shi, Zhen Xia, Jiajing Zhang, Robert Best, Chuanjie Wu, Jay W Ponder, and Pengyu Ren. Polarizable atomic multipole-based amoeba force field for proteins. *Journal of chemical theory and computation*, 9(9):4046–4063, 2013.

[68] Zhifeng Jing, Chengwen Liu, Sara Y Cheng, Rui Qi, Brandon D Walker, Jean-Philip Piquemal, and Pengyu Ren. Polarizable force fields for biomolecular simulations: Recent advances and applications. *Annual Review of biophysics*, 48(1):371–394, 2019.

[69] Louis Lagardère, Luc-Henri Jolly, Filippo Lipparini, Félix Aviat, Benjamin Stamm, Zhifeng F Jing, Matthew Harger, Hedieh Torabifard, G Andrés Cisneros, Michael J Schnieders, et al. Tinker-hp: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chemical science*, 9(4):956–972, 2018.

[70] Pedro E. M. Lopes, Jing Huang, Jihyun Shim, Yun Luo, Hui Li, Benoît Roux, and Alexander D. Jr. MacKerell. Polarizable force field for peptides and proteins based on the classical drude oscillator. *Journal of Chemical Theory and Computation*, 9(12):5430–5449, 2013.

[71] Fang-Yu Lin and Alexander D MacKerell Jr. Force fields for small molecules. In *Biomolecular simulations: Methods and protocols*, pages 21–54. Springer, 2019.

[72] Julian D. Gale, Luc M. LeBlanc, Peter R. Spackman, Alessandro Silvestri, and Paolo Raiteri. A Universal Force Field for Materials, Periodic GFN-FF: Im-

plementation and Examination. *J. Chem. Theory Comput.*, 17(12):7827–7849, 2021.

[73] M. Aschi, R. Spezia, A. Di Nola, and A. Amadei. A first-principles method to model perturbed electronic wavefunctions: the effect of an external homogeneous electric field. *Chem. Phys. Lett.*, 344(3-4):374–380, 2001.

[74] Hai Lin and Donald G. Truhlar. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.*, 117(2):185, 2007.

[75] Abir Ganguly, Eliot Boulanger, and Walter Thiel. Importance of mm polarization in qm/mm studies of enzymatic reactions: Assessment of the qm/mm drude oscillator model. *Journal of Chemical Theory and Computation*, 13(6):2954–2961, 2017.

[76] Qiang Cui, Tanmoy Pal, and Luke Xie. Biomolecular qm/mm simulations: What are some of the "burning issues"? *The Journal of Physical Chemistry B*, 125(3):689–702, 2021.

[77] Mats Svensson, Stéphane Humbel, Robert D. J. Froese, Toshiaki Matsubara, Stefan Sieber, and Keiji Morokuma. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels–Alder Reactions and $Pt(P(t\text{-}Bu)_3)_2 + H_2$ Oxidative Addition. *J. Phys. Chem.*, 100(50):19357–19363, 1996.

[78] Feliu Maseras and Keiji Morokuma. Imomm: A new integrated ab-initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.*, 16(9):1170–1179, 1995.

[79] Thom Vreven, K. Suzie Byun, István Komáromi, Stefan Dapprich, John A. Jr. Montgomery, Keiji Morokuma, and Michael J. Frisch. Combining quantum mechanics methods with molecular mechanics methods in oniom. *Journal of Chemical Theory and Computation*, 2(3):815–826, 2006.

[80] Alexander Zlobin, Julia Belyaeva, and Andrey Golovin. Challenges in Protein QM/MM Simulations with Intra-Backbone Link Atoms. *J. Chem. Inf. Model.*, 63(2):546–560, 2023.

[81] Laura Zanetti-Polzi, Dimitrios A. Pantazis, and Isabella Daidone. Intermolecular Photoinduced Electron Transfer in Biosystems: Impact of Conformational Transitions and Multiple Channels on Kinetics. *ChemPhotoChem*, 2024.

[82] Mats H. M. Olsson, Chresten R. Søndergaard, Michal Rostkowski, and Jan H. Jensen. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.*, 7(2):525–537, 2011.

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

[84] Fu Lin and Renxiao Wang. Systematic Derivation of AMBER Force Field Parameters Applicable to Zinc-Containing Systems. *J. Chem. Theory Comput.*, 6(6):1852–1870, 2010.

[85] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, 1983.

[86] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: A NlogN method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[87] J.C. Berendsen et al. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

[88] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1), 2007.

[89] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

[90] R. Kumar, J. R. Schmidt, and J. L. Skinner. Hydrogen bonding definitions and dynamics in liquid water. *J. Chem. Phys.*, 126(20), 2007.

[91] Julien Guénolé, Wolfram G. Nöhring, Aviral Vaid, Frédéric Houllé, Zhuocheng Xie, Aruna Prakash, and Erik Bitzek. Assessment and optimization of the fast inertial relaxation engine (fire) for energy minimization in atomistic simulations and its implementation in lammps. *Comput. Mater. Sci.*, 175:109584, 2020.

[92] Alexey V. Onufriev and David A. Case. Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.*, 48(1):275–296, 2019.

[93] Axel D. Becke. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.*, 98(2):1372–1377, 1993.

[94] P. C. Hariharan and J. A. Pople. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta*, 28(3):213–222, 1973.

[95] 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. *J. Comput. Chem.*, 31(4):671–690, 2010.

[96] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L de Groot, Helmut Grubmüller, and Alexander D MacKerell. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods*, 14(1):71–73, 2017.

[97] Damien Sorigué, Bertrand Légeret, Stéphan Cuiné, Stéphanie Blangy, Solène Moulin, Emmanuelle Billon, Pierre Richaud, Sabine Brugière, Yohann Couté, Di-

dier Nurizzo, Pavel Müller, Klaus Brettel, David Pignol, Pascal Arnoux, Yonghua Li-Beisson, Gilles Peltier, and Fred Beisson. An algal photoenzyme converts fatty acids to hydrocarbons. *Science*, 357(6354):903–907, 2017.

[98] William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, 1985.

[99] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52(12):7182–7190, 1981.

[100] Frank Neese. The ORCA program system. *WIREs Comput. Mol. Sci.*, 2(1):73–78, 2012.

[101] Frank Neese, Frank Wennmohs, Ute Becker, and Christoph Riplinger. The ORCA quantum chemistry program package. *J. Chem. Phys.*, 152(22), 2020.

[102] Christine E. Schulz, Maurice van Gastel, Dimitrios A. Pantazis, and Frank Neese. Converged Structural and Spectroscopic Properties for Refined QM/MM Models of Azurin. *Inorg. Chem.*, 60(10):7399–7412, 2021.

[103] Mahesh Sundararajan, Christoph Riplinger, Maylis Orio, Frank Wennmohs, and Frank Neese. Spectroscopic Properties of Protein-Bound Cofactors: Calculation by Combined Quantum Mechanical/Molecular Mechanical (QM/MM) Approaches. In *Encycl. Inorg. Bioinorg. Chem.* Wiley, 2011.

[104] Manuel Hitzenberger and Thomas S. Hofer. Probing the range of applicability of structure- and energy-adjusted QM/MM link bonds. *J. Comput. Chem.*, 36(26):1929–1939, 2015.

[105] Takeshi Yanai, David P Tew, and Nicholas C Handy. A new hybrid exchange–correlation functional using the coulomb-attenuating method (cam-b3lyp). *Chemical Physics Letters*, 393(1):51–57, 2004.

[106] Achintya Kumar Dutta, Frank Neese, and Róbert Izsák. Speeding up equation of motion coupled cluster theory with the chain of spheres approximation. *J.*

Chem. Phys., 144(3), 2016.

[107] R. Fletcher. *Practical Methods of Optimization*. Wiley, 2000.

[108] Frank Neese. Software update: The orca program system—version 6.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 15(2):e70019, 2025.

[109] So Hirata and Martin Head-Gordon. Time-dependent density functional theory within the Tamm–Dancoff approximation. *Chem. Phys. Lett.*, 314(3-4):291–299, 1999.

[110] Marco D'Abramo, Massimiliano Aschi, and Andrea Amadei. Theoretical modeling of UV-Vis absorption and emission spectra in liquid state systems including vibrational and conformational effects: Explicit treatment of the vibronic transitions. *J. Chem. Phys.*, 140(16), 2014.

[111] Qiang Cui, Marcus Elstner, and Martin Karplus. A Theoretical Analysis of the Proton and Hydride Transfer in Liver Alcohol Dehydrogenase (LADH). *J. Phys. Chem. B*, 106(10):2721–2740, 2002.

[112] Simon P. Webb, Pratul K. Agarwal, and Sharon Hammes-Schiffer. Combining Electronic Structure Methods with the Calculation of Hydrogen Vibrational Wavefunctions: Application to Hydride Transfer in Liver Alcohol Dehydrogenase. *J. Phys. Chem. B*, 104(37):8884–8894, 2000.

[113] Bryce V. Plapp, Baskar Raj Savarimuthu, Daniel J. Ferraro, Jon K. Rubach, Eric N. Brown, and S. Ramaswamy. Horse Liver Alcohol Dehydrogenase: Zinc Coordination and Catalysis. *Biochemistry*, 56(28):3632–3646, 2017.

[114] Laurie A. LeBrun, Doo-Hong Park, S. Ramaswamy, and Bryce V. Plapp. Participation of histidine-51 in catalysis by horse liver alcohol dehydrogenase,. *Biochemistry*, 43(11):3014–3026, 2004.

[115] Kang Man Lee and Su Yeon Son. Determination of the pka for histidine-51 residue in the ternary complex of horse liver alcohol dehydrogenase. *Archives of*

Pharmacal Research, 15(3):229–233, 1992.

[116] D. Sorigué, K. Hadjidemetriou, S. Blangy, G. Gotthard, A. Bonvalet, N. Coquelle, P. Samire, A. Aleksandrov, L. Antonucci, A. Benachir, S. Boutet, M. Byrdin, M. Cammarata, S. Carbajo, S. Cuiné, R. B. Doak, L. Foucar, A. Gorel, M. Grünbein, E. Hartmann, R. Hienerwadel, M. Hilpert, M. Kloos, T. J. Lane, B. Légeret, P. Legrand, Y. Li-Beisson, S. L. Y. Moulin, D. Nurizzo, G. Peltier, G. Schirò, R. L. Shoeman, M. Sliwa, X. Solinas, B. Zhuang, T. R. M. Barends, J.-P. Colletier, M. Joffre, A. Royant, C. Berthomieu, M. Weik, T. Domratcheva, K. Brettel, M. H. Vos, I. Schlichting, P. Arnoux, P. Müller, and F. Beisson. Mechanism and dynamics of fatty acid photodecarboxylase. *Science*, 372(6538), 2021.

[117] Derren J. Heyes, Balaji Lakavath, Samantha J. O. Hardman, Michiyo Sakuma, Tobias M. Hedison, and Nigel S. Scrutton. Photochemical Mechanism of Light-Driven Fatty Acid Photodecarboxylase. *ACS Catal.*, 10(12):6691–6696, 2020.

[118] Ruiqi Wu, Xiankun Li, Lijuan Wang, and Dongping Zhong. Ultrafast Dynamics and Catalytic Mechanism of Fatty Acid Photodecarboxylase. *Angew. Chemie Int. Ed.*, 61(50), 2022.

[119] Koji Hasegawa, Shinji Masuda, and Taka-aki Ono. Light Induced Structural Changes of a Full-length Protein and Its BLUF Domain in YcgF(Blrp), a Blue-Light Sensing Protein That Uses FAD (BLUF). *Biochemistry*, 45(11):3785–3793, 2006.

[120] Zheyun Liu, Meng Zhang, Xunmin Guo, Chuang Tan, Jiang Li, Lijuan Wang, Aziz Sancar, and Dongping Zhong. Dynamic determination of the functional state in photolyase and the implication for cryptochrome. *Proc. Natl. Acad. Sci.*, 110(32):12972–12977, 2013.

[121] M. Sikorski, E. Sikorska, A. Koziolowa, R. Gonzalez Moreno, J.L. Bourdelande, R.P. Steer, and F. Wilkinson. Photophysical properties of lumichromes in water.

J. Photochem. Photobiol. B Biol., 60(2-3):114–119, 2001.

[122] F. Plasser. Theodore: A toolbox for a detailed and automated analysis of electronic excited state computations. *The Journal of Chemical Physics*, 152(8):084108, 2020.

[123] Richard L. Martin. Natural transition orbitals. *The Journal of Chemical Physics*, 118(11):4775–4777, 2003.

[124] Gianluca Dell'Orletta, Nico Di Fonte, Matteo Farina, and Isabella Daidone. Insights into substrate protonation and solvent accessibility in the active site of fatty acid photodecarboxylase. *The Journal of Physical Chemistry Letters*, 0(0):12538–12544, 0. PMID: 41296541.

[125] Mai Van Bay, Nguyen Khoa Hien, Phan Thi Diem Tran, Nguyen Tran Kim Tuyen, Doan Thi Yen Oanh, Pham Cam Nam, and Duong Tuan Quang. TD-DFT benchmark for UV-Vis spectra of coumarin derivatives. *Vietnam J. Chem.*, 59(2):203–210, 2021.

[126] Péter Pál Fehér, Ádám Madarász, and András Stirling. A Practice-Oriented Benchmark Strategy to Predict the UV-Vis Spectra of Organic Photocatalysts . *Chemistry–Methods*, 3(5), 2023.

[127] Costantino Zazza, Andrea Amadei, Nico Sanna, Andrea Grandi, Giovanni Chillemi, Alfredo Di Nola, Marco D'Abramo, and Massimiliano Aschi. Theoretical modeling of the valence UV spectra of 1,2,3-triazine and uracil in solution. *Phys. Chem. Chem. Phys.*, 8(12):1385, 2006.

[128] Amit Manor Armon, Anjan Bedi, Veniamin Borin, Igor Schapiro, and Ori Gidron. Bending versus twisting acenes – a computational study. *European Journal of Organic Chemistry*, 2021(39):5424–5429, 2021.

[129] Mohammad Pabel Kabir, Daniel Ouedraogo, Yoelvis Orozco-Gonzalez, Giovanni Gadda, and Samer Gozem. Alternative Strategy for Spectral Tuning of Flavin-Binding Fluorescent Proteins. *J. Phys. Chem. B*, 127(6):1301–1311, 2023.

[130] Alexey Aleksandrov, Adeline Bonvalet, Pavel Müller, Damien Sorigué, Fred Beisson, Laura Antonucci, Xavier Solinas, Manuel Joffre, and Marten H. Vos. Catalytic Mechanism of Fatty Acid Photodecarboxylase: On the Detection and Stability of the Initial Carbonyloxy Radical Intermediate. *Angew. Chemie Int. Ed.*, 63(19), 2024.

[131] Robin D. Rogers and Kenneth R. Seddon. Ionic liquids–solvents of the future? *Science*, 302(5646):792–793, 2003.

[132] Tom Welton. Ionic liquids: a brief history. *Biophysical reviews*, 10(3):691–706, 2018.

[133] Matteo Farina, Flaminia Rondino, Andrea Lapi, Mauro Falconieri, Serena Gagliardi, Isabella Daidone, Caterina Fraschetti, Enrico Bodo, and Antonello Filippi. Applying computational spectroscopy methods to raman spectra of dicationic, imidazolium-based, ionic liquids. *J. of Phys. Chem. B*, 128(43):10650–10660, 2024.

[134] Andrzej J. Sadlej. Medium-size polarized basis sets for high-level correlated calculations of molecular electric properties. *Collect. Czechoslov. Chem. Commun.*, 53(9):1995–2016, 1988.

[135] José N. Canongia Lopes, Johnny Deschamps, and Agílio A. H. Pádua. Modeling Ionic Liquids Using a Systematic All-Atom Force Field. *J. Phys. Chem. B*, 108(6):2038–2047, 2004.

[136] Andreia S. L. Gouveia, Carlos E. S. Bernardes, Liliana C. Tomé, Elena I. Lozinskaya, Yakov S. Vygodskii, Alexander S. Shaplov, José N. Canongia Lopes, and Isabel M. Marrucho. Ionic liquids with anions based on fluorosulfonyl derivatives: from asymmetrical substitutions to a consistent force field model. *Phys. Chem. Chem. Phys.*, 19(43):29617–29624, 2017.

[137] E. Bodo, M. Chiricotto, and R. Caminiti. Structure of Geminal Imidazolium Bis(trifluoromethylsulfonyl)imide Dicationic Ionic Liquids: A Theoretical Study

of the Liquid Phase. *J. Phys. Chem. B*, 115(49):14341–14347, 2011.

[138] Alessandra Serva, Valentina Migliorati, Andrea Lapi, Giuliana Aquilanti, Alessandro Arcovito, and Paola D'Angelo. Structural properties of geminal dicationic ionic liquid/water mixtures: a theoretical and experimental insight. *Phys. Chem. Chem. Phys.*, 18(24):16544–16554, 2016.

[139] Andrea Amadei, Antonius B. M. Linssen, and Herman J. C. Berendsen. Essential dynamics of proteins. *Proteins Struct. Funct. Bioinforma.*, 17(4):412–425, 1993.

[140] Isabella Daidone and Andrea Amadei. Essential dynamics: foundation and applications. *WIREs Comput. Mol. Sci.*, 2(5):762–770, 2012.

[141] Marta Monti, Mauro Stener, and Massimiliano Aschi. A computational approach for modeling electronic circular dichroism of solvated chromophores. *J. Comput. Chem.*, 43(30):2023–2036, 2022.

[142] Vitor H. Paschoal, Luiz F. O. Faria, and Mauro C. C. Ribeiro. Vibrational spectroscopy of ionic liquids. *Chem. Rev.*, 117(10):7053–7112, 2017.

[143] Zahra Ostadsharif Memar and Majid Moosavi. Assessing OPLS-based force fields for investigating the characteristics of imidazolium-based dicationic ionic liquids: A comparative study with AIMD simulations and experimental findings. *J. Chem. Phys.*, 159(24), 2023.

[144] José N. Canongia Lopes, Karina Shimizu, Agílio A. H. Pádua, Yasuhiro Umebayashi, Shuhei Fukuda, Kenta Fujii, and Shin-ichi Ishiguro. A Tale of Two Ions: The Conformational Landscapes of Bis(trifluoromethanesulfonyl)amide and *N,N*-Dialkylpyrrolidinium. *J. Phys. Chem. B*, (5):1465–1472.

[145] J. C. Lassègues, J. Grondin, R. Holomb, and P. Johansson. Raman and ab-initio study of the conformational isomerism in the 1-ethyl-3-methyl-imidazolium bis(trifluoromethanesulfonyl)imide ionic liquid. *J. Raman Spectrosc.*, 38(5):551–558, 2007.

[146] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.*, 10(44):6615, 2008.

[147] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of Chemical Physics*, 132(15):154104, 04 2010.

[148] G. Arfken. *The Method of Steepest Descents*. Orlando, 1985.

[149] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. S15 M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. . GAUSSIAN 16, 2016.

[150] Rolf W. Berg. Raman Spectroscopy and Ab-Initio Model Calculations on Ionic Liquids. *Monatshefte für Chemie - Chem. Mon.*, 138(11):1045–1075, 2007.

[151] Matteo Turisini, Mirko Cestari, and Giorgio Amati. LEONARDO. *J. large-scale Res. Facil. JLSRF*, 9(1), 2024.

5.6 Copyright Note