U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

PhD

3.º
CICLO

FCUP
UNL
UA
2018

U.PORTO

The origin of the catalytic power of enzymes

Ana Rita de Almeida Calixto Silva

FC

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

UNIVERSIDADE NOVA DE LISBOA

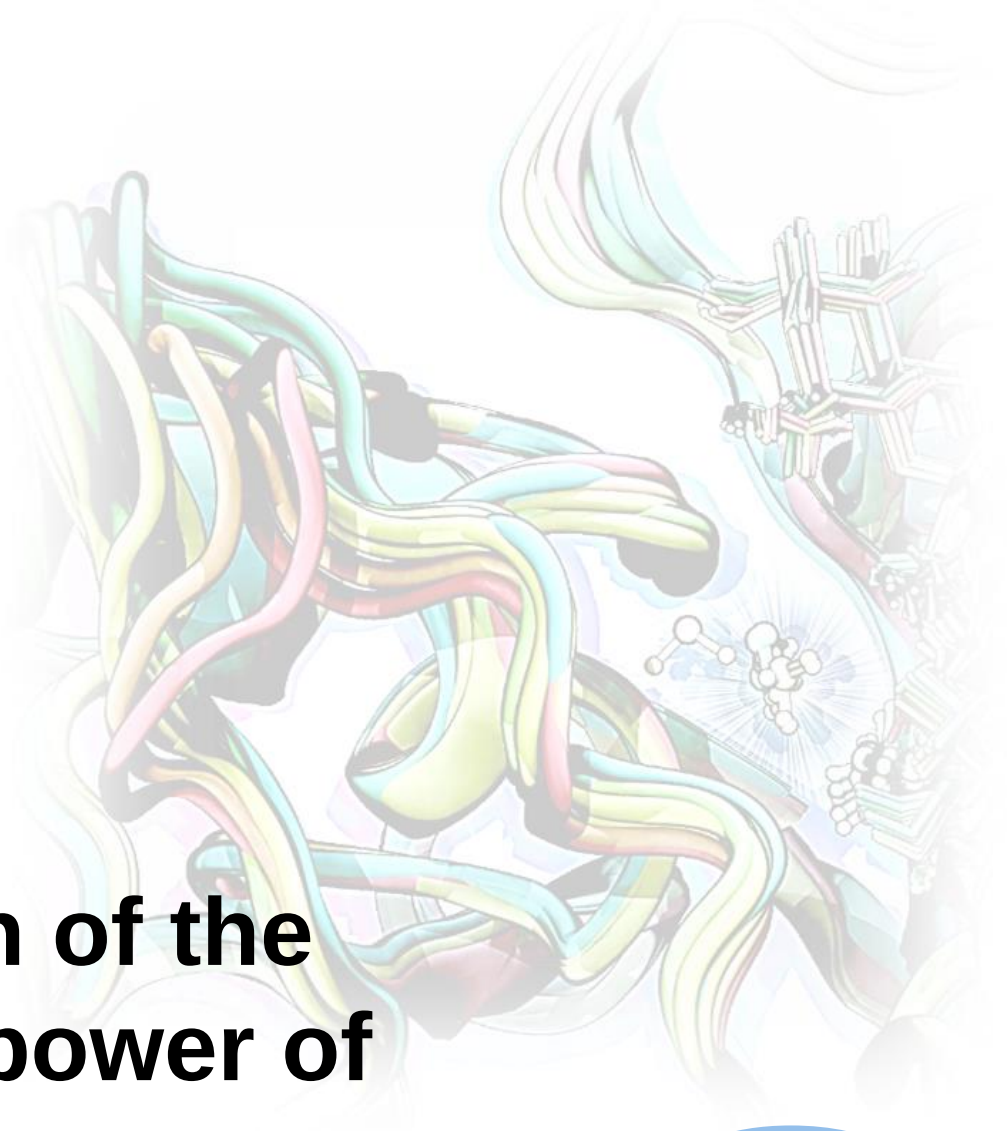
universidade de aveiro

The origin of the catalytic power of enzymes

Ana Rita de Almeida Calixto Silva

Tese de Doutoramento apresentada à
Faculdade de Ciências da Universidade do Porto, Universidade Nova de Lisboa e Universidade de Aveiro
Química

2018



The origin of the catalytic power of enzymes

Ana Rita de Almeida Calixto Silva

Programa Doutoral em Química Sustentável

Departamento de Química e Bioquímica

2018

Orientador

Pedro Alexandrino Fernandes, Professor Associado, FCUP

Coorientador

Maria João Ramos, Professora Catedrática, FCUP

Acknowledgements

First of all, I would like to thank Professor Pedro Alexandrino Fernandes, my supervisor, for giving me the opportunity to work with him, for the continuous support, helpful suggestions, motivation and enthusiastic guidance, even when the results were not as good as we expected. I am also grateful to Professor Maria João Ramos, my co-supervisor, for the opportunity to work in her group for all these years and for having always some space in her busy schedule for fruitful discussions and valuable suggestions.

I would like to acknowledge all the current and past colleagues of the Theoretical and Computational Biochemistry group. Thank you for providing such wonderful work environment, for all the productive discussions during our internal seminars, for your patience, enthusiasm, support and for the fun we had together. A special thanks to Natércia Brás for sharing her knowledge and experience with me, during my first years in the group. I would like to thank Pedro Ferreira, for his fast learning skills and for his precious help to finish Natércia's work. I am also thankful to Diogo Martins and António Ribeiro for all the productive suggestions and support with some works of this thesis. To Cátia Moreira, Diana Gesto, Fabiola Medina, João Coimbra and Rui Neves a special thank, not only for all your helpful suggestions to improve my work, but also for all your friendship during the last years.

I am especially grateful to my mother, my brother and Ri, for their everyday support and unconditional love, that were essential for the progress of this thesis.

Finally, I would like to thank Fundação para a Ciência e Tecnologia for my PhD grant SFRH/BD/95962/2013.

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

Cofinanciado por:
PORTUGAL 2020
UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional
NORTE 2020

“The important thing is not to stop questioning. Curiosity has its own reason for existing.

Albert Einstein

Abstract

The main goal of this thesis was to understand the origin of the catalytic power of enzymes. Enzymes are fundamental to life, being able to process their physiological substrates with astonishing kinetics. During the last years, many different theories have been proposed to explain how they do this. However, all of them are undemonstrated and not consensual. Here, we tried to produce some insights in this controversial field, employed a combination of theoretical and computational methods, in different enzymes, to try to describe their catalytic mechanisms, or even to reveal some important keys into the question of the origin of their catalytic power.

A brief outline of enzymes and some possible hypothesis to explain enzyme catalysis is given in **Chapter 1**. **Chapter 2** consists on an overview of the computational methods used throughout the works presented in this thesis. The remaining chapters concern the work developed during the last four years. These chapters are not presented chronologically.

Chapter 3 consists in a systematic analysis of three parameters directly related with enzyme catalysis (activation free energy, substrate binding and enzyme efficiency) and their correlation with structural and biological information, such as, for example, the presence of cofactors, their size, oligomerization state and cellular location. Comparisons between these correlations in all classes of enzymes, showed that, regardless of the large diversity of these biomolecules, the evaluated parameters fall in a very narrow range of values.

The next two chapters correspond to mechanistic studies on two different enzymes, Human Renin (**Chapter 4**) and PatG macrocyclase (**Chapter 5**), using QM/MM methodologies.

Chapter 6 consists in a more methodological work. In this study the influence of fixing residues during a QM/MM study was explored and the results showed that, this widely used approximation, is safe if the model has, at least, a region of 6.00Å of free residues around the active site.

The two final chapters addressed the origin of the catalytic power of enzymes, exploring the influence of enzyme conformation and flexibility on the reaction barriers. In **Chapter 7**, different conformations of human pancreatic alpha-amylase, were sampled and the energy profiles along the reaction coordinates were explored. We found that the position and orientation of a buried, non-reactive, water molecule influences the reactivity of this enzyme on a sub-nanosecond timescale. In **Chapter 8** the influence of enzyme conformations was studied on the catalytic mechanism of HIV-1 protease. In a similar way

to the previous work, the catalytic mechanism of this enzyme was explored starting from different initial structures. The results showed that small variations on the orientation of the active site residues leads to differences in the progress of the reaction and in the activation barriers.

Although there are different and non-correlated works along this thesis, there are a unifying topic that is common to all of them: enzyme catalysis. The results presented in the next chapters provide, not only, a better understand on the catalytic mechanism of different enzymes (**Chapter 4** and **Chapter 5**), but also some important hints on the origin of the catalytic power of enzymes, showing that the conformational fluctuations seem to influence the catalysis (**Chapter 7** and **Chapter 8**).

Keywords: Enzymes, Enzyme Catalysis, Catalytic Power, Computational Biochemistry, Molecular Modelling, Hybrid QM/MM Methodologies, ONIOM.

Resumo

Esta tese tinha como objetivo principal o estudo e compreensão da origem do poder catalítico das enzimas. As enzimas são biomoléculas essenciais capazes de processar os seus substratos naturais com uma cinética surpreendente e compatível com a vida.
Durante os últimos anos, diferentes teorias têm sido propostas para tentar explicar de que forma estas biomoléculas são capazes de catalisar tais reações. Contudo, até agora, nenhuma delas foi verdadeiramente demonstrada, sendo este um tema ainda bastante controverso. Com os trabalhos apresentados ao longo desta tese, tentamos dar algum contributo para uma melhor compreensão da origem do poder catalítico das enzimas. Para isso, foram utilizados métodos teóricos e computacionais, em diferentes enzimas, de forma a descrever os mecanismos pelos quais estas catalisam as suas reações, ou, em última instância, de forma a contribuir com importantes conceitos que ajudem a dar resposta à questão em aberto desta tese: qual a origem do poder catalítico das enzimas?
Uma breve apresentação das enzimas como catalisadores biológicos e de algumas das hipóteses propostas para explicar a origem do seu poder catalítico, são apresentados no **Capítulo 1**. No **Capítulo 2**, é dada uma visão geral dos métodos computacionais utilizados para realizar cada um dos trabalhos realizados.
Os capítulos seguintes dizem respeito aos trabalhos desenvolvidos ao longo dos últimos 4 anos. Os mesmos não se encontram apresentados cronologicamente por uma questão de organização de conceitos.
O **Capítulo 3** consiste numa análise sistemática de três parâmetros diretamente relacionados com a catálise enzimática (energia livre de ativação, ligação do substrato e eficiência enzimática) e a sua relação com informações estruturais e biológicas disponíveis para cada classe de enzimas. Foram analisados estes parâmetros em enzimas com e sem cofatores, com diferentes tamanhos e estados de oligomerização, diferente localização celular, entre outros fatores. Os resultados obtidos mostraram que apesar da grande diversidade de enzimas conhecidas, os parâmetros avaliados são bastante semelhantes entre todas elas, independentemente da classe a que pertencem.
Os dois capítulos seguintes correspondem a estudos mecanísticos de duas enzimas diferentes, a renina humana (**Capítulo 4**) e a PatG macrociclase (**Chapter 5**), usando métodos QM/MM.
O **Capítulo 6** descreve um estudo metodológico, no qual foi avaliada a influência do congelamento de resíduos, durante os estudos de mecanismos enzimáticos usando métodos QM/MM, na barreira da reação estudada. Os resultados obtidos mostraram que esta aproximação, largamente usada neste tipo de estudos, é segura, desde que o

sistema esteja livre numa região de, pelo menos, 6Å em redor dos resíduos que participam diretamente na reação.

Os dois capítulos finais, abordam mais diretamente a questão da origem do poder catalítico das enzimas, explorando a influência de diferentes conformações da enzima nas barreiras de ativação da reação catalisada. No **Capítulo 7**, diferentes conformações da enzima alfa-amilase (pancreática humana) foram utilizadas para estudar o perfil de energia livre ao longo da coordenada da reação. Os resultados mostraram que a posição e orientação de uma molécula de água, enterrada no centro ativo, influência a reatividade desta enzima, numa escala temporal na ordem do nanossegundo. No **Capítulo 8**, a influência de diferentes conformações da enzima no mecanismo catalítico da protease do HIV-1 foi também avaliada. À semelhança do trabalho anterior o mecanismo desta enzima foi estudado partindo de diferentes estruturas iniciais. Os resultados obtidos mostraram que a orientação dos resíduos do centro ativo influência a progressão e as barreiras da reação em estudo.

Apesar desta tese conter trabalhos que não estão diretamente ligados entre si, todos eles possuem um tópico em comum: a catálise enzimática. Os resultados apresentados dão, não só a conhecer os detalhes atomísticos do mecanismo catalítico de diferentes enzimas (**Capítulos 4 e 5**), como também fornecem pistas importantes sobre a origem do poder catalítico das enzimas, mostrando que flutuações conformacionais podem influenciar a catálise enzimática (**Capítulos 7 e 8**).

Palavras-chave: Enzimas, Catálise Enzimática, Poder Catalítico, Bioquímica Computacional, Modelação Molecular, Métodos Híbridos QM/MM, ONIOM.

List of Abbreviations

AMBER	Assisted Model Building with Energy Refinement
BRENDA	Braunschweig Enzyme Database
CGTO	Contracted Gaussian Type Orbital
DFT	Density Functional Theory
E	Enzyme
E.C	Enzyme Commission number
ES	Enzyme-Substrate complex
EE	Electrostatic Embedding
EVB	Empirical Valence Bond
GGA	Generalized Gradient Approximation
GTO	Gaussian Type Orbital
HF	Hatree Fock
HIV	Human Immunodeficiency Virus
INT	Reaction Intermediate
IRC	Intrinsic Reaction Coordinates
K_{cat}	First order rate constant
K_M	Michaelis-Menten constant
LDA	Local Density Approximation
LHB	Low Barrier Hydrogen Bonds
MD	Molecular Dynamics
ME	Mechanical Embedding
MM	Molecular Mechanics
NACs	Near Attack Conformations
NPT	Isobaric Isothermal Ensemble
NVT	Canonical Ensemble
ONIOM	Our own N-layered Integrated molecular Orbital and molecular Mechanics
P	Product
PBC	Periodic Boundary Conditions
PDB	Protein Data Bank
PES	Potential Energy Surface
PR	Protease
PGTO	Primitive Gaussian Type Orbital
PME	Particle Mesh-Ewald
QM	Quantum Mechanics

QM/MM	Quantum Mechanics / Molecular Mechanics
RESP	Restrained Electrostatic Potential
RMSd	Root Mean Square Deviation
SCF	Self Consistent Field
TIP3P	Transferable Intermolecular Potential 3P
TS	Transition State
TST	Transition State Theory
ZPE	Zero Point Energy

List of Publications

This thesis is based on following published works and in two other works in progress

- Calixto, A. R.; Bras, N. F.; Fernandes, P. A. and Ramos, M. J., ***Reaction Mechanism of Human Renin Studied by Quantum Mechanics/Molecular Mechanics (QM/MM) Calculations***. ACS Catalysis, 2014, 4 (11), 3869-3876 **DOI: 10.1021/cs500497f7**

- Bras, N. F.; Ferreira, P.; **Calixto, A. R.;** Jaspars, M.; Houssen, W.; Naismith, J. H.; Fernandes, P. A and Ramos, M. J., ***The Catalytic Mechanism of the Marine-Derived Macrocyclase PatGmac***. Chemistry-European Journal, 2016, 22 (37), 13089-13097 **DOI: 10.1002/chem.201601670**

- Calixto, A. R.; Ramos, M. J. and Fernandes, P. A., ***Influence of Frozen Residues on the Exploration of the PES of Enzyme Reaction Mechanisms***. Journal of Chemical Theory and Computation, 2017, 13 (11), 5486-5495 **DOI: 10.1021/acs.jctc.7b00768**

- Santos-Martins, D.; **Calixto, A. R**.; Fernandes, P. A. and Ramos M. J., ***Water Controls Reactivity in Alpha-amylase on a Subnanosecond Timescale***. ACS Catalysis, 2018, 8, 4055-4063 DOI: 10.1021/acscatal.7b04400

- Sousa, S. F.; **Calixto, A. R**.; Ramos, M. J.; Lim C. and Fernandes P.; ***Activation Free Energy, Substrate Binding Free Energy and Enzyme Efficiency Fall in a Very Narrow Range of values for all enzymes***. (Manuscript in preparation)

Calixto, A. R.; Ramos, M. J. and Fernandes P.; ***The influence of enzyme conformation on the HIV-1 protease catalytic mechanism*** (Manuscript in preparation)

Index

Index of Figures

Index of Tables

CHAPTER 1. Enzymes as life catalysts

It was obvious—to me at any rate—that the answer was to why an enzyme is able to speed up a chemical reaction by as much as 10 million times. It had to do this by lowering the energy of activation—the energy of forming the activated complex. It could do this by forming strong bonds with the activated complex, but only weak bonds with the reactants or products.

Linus Pauling

1.1 Enzymes: An introduction

Enzymes are fundamental for life. They are proteins with exceptional molecular devices that are able to determine the patterns of biochemical transformations. Their catalytic power and specificity are their most remarkable characteristics. Enzymes are able to bind a very large number of different molecules and capable to catalyze a vastly number of different reactions.

Why these special proteins are so important for life? Enzymes are essential for life, due to their responsibility for almost all biological reactions. In the absence of these proteins, some of those reactions are among the slowest that have ever been measure, being incompatible with life [1,2].

The large efficiency of enzymes could be demonstrated by innumerable examples. One of the most common, giving by biochemistry books [3,4], is the hydration of carbon dioxide by the enzyme carbonic anhydrase, which allows a complete and fast transference of this molecule from the tissues, into the blood and to the alveolus. This vital reaction is 10^7 times as fast as the uncatalyzed one.

The catalytic power of enzymes is much larger than the one observed for synthetic or inorganic catalysts. In addition, they work in aqueous solution at mild temperatures and pH, in contrast with non-biological catalysts, which need very drastic conditions to accelerate chemical reactions.

Studies on these life catalysts have a very important role, not only to improve our knowledge of the biologic systems, but also to understand and treat many diseases. Many drugs can exert their effect through chemical interactions with enzymes. In addition to its important role in medicine, enzymes can also improve other sectors as chemical and food industry, being alternatives to the synthetic catalysts.

1.1.1 Historical background

The history of enzymes started in the late 1700's, with studies that described the digestion of food by secretions of the stomach. More than a century after, Louis Pasteur concluded that yeasts have "ferments" that were responsible for the digestion of food [5]. These "ferments", which were inseparable from the living form of yeasts cells, were known as "vitalism". This old concept of "vitalism" evolved, years later, for the concept of enzymes, which was first used by Frederick W. Kühne [6]. Some years later (1926) with the isolation and crystallization of urease, enzymes were considered as proteins for the first time [7]. However, this idea has only been confirmed some years after, with the crystallization of pepsin, trypsin and other digestive proteins [8,9].

Since these discoveries, the molecular nature of enzymes and how they catalyze these reactions have been questioned. The first theory on the origin of the catalytic power of enzymes were proposed by J.B.S. Haldane and suggested that weak bonding interactions between enzymes and its substrates might explain the acceleration of the reactions that they catalyze [10]. Some years later Pauling, with new insights about protein structure, reflects about the Haldane's idea of enzyme action and wrote *"…the only reasonable picture of the catalytic activity of enzymes is that which involves an active region of the surface of the enzyme which is closely complementary in structure not to the substrate molecule itself in its normal configuration, but rather to the substrate molecule in a strained conformation corresponding to the 'activated complex' for the reaction catalyzed by the enzyme".* He also wrote the sentence, by which we started this chapter [11], formulating one of the first interpretation assuming the complementarity between the enzyme's active site and the transition state structure.

Since then, the research on enzymes has grown quickly and the structure of thousands of enzymes has been elucidated, as well as their catalytic mechanism.

The next sections will explore the classification of different classes of enzymes and the principles underlying their efficiency. We will end this chapter with a discussion on the different actual literature suggestions that try to explain the origin of the catalytic power of enzymes.

1.2 Enzymes classification

Enzymes are classified by the reactions they catalyze. Usually the suffix “ase” is common for many enzymes and, it is added to the name of the substrate or to a word that describes the enzyme’s function. For example, HIV-1 prote***ase*** (studied in this thesis) is an enzyme from HIV-1 virus which catalyze the hydrolysis of essential proteins for virus replication. α-amyl***ase*** (also studied here) catalysis the hydrolysis of alpha bonds of alpha-linked polysaccharides.

Other enzymes follow a different nomenclature, given by them discovers. For example, the name of the enzyme “renin” (also studied in this thesis), an enzyme secreted by the kidneys, comes from *ren* + *in* (“*kidney*” + “*compound*”). This enzyme is also known as angiotensinogen***ase***, because it cleaves the substrate angiotensinogen.

In this case, and in many others, the enzyme has two names, which could difficult a uniform classification. To avoid ambiguities, and because of the increasing of newly enzymes, an international agreement was created to classify enzymes by a unique name. This system divides enzymes into six different classes, based on the type of reaction that they catalyze: **Oxidoreductases**, **Transferases**, **Hydrolases**, **Lyases**, **Isomerases** and **Ligases**. Each class has, in its turn, different subclasses. These international classification of enzymes assigns each enzyme to a code of four numbers (Enzyme Commission number, E.C. number) and a systematic name, which identify the reaction it catalyzes. For example, the E.C. number of the enzyme renin is 3.4.23.15. The first number of the code denotes the enzyme class (Class 3 - Hydrolases), the second number, the subclass (Subclass 4 - acting on peptide bonds (peptidases)), the third number (23), indicates that it is an aspartic endopeptidase and the last number (15) is unique for renin. In the **Chapter 3** of this thesis, we will analyze different parameters related with enzyme catalysis for each class of enzymes and, because of that, we present a brief description of each class below. **Table 1.1** summarizes the type of reaction catalysed by them.

E.C. 1: Oxidoreductases

Oxidoreductases are a class of enzymes that catalyze transfer of electrons from one molecule (reductant) to an acceptor molecule (oxidant). These enzymes catalyze

reactions that follows the general scheme (for a single electron transfer): $A^n + B^m \rightarrow A^{n+1} + B^{m-1}$, where A is the reductant and B the oxidant. Their systematic name is usually *donor:acceptor oxidoreductase*. The common name of enzymes of this group is *dehydrogenase*. Sometimes, *reductase* is used as an alternative. *Oxidase* is only used .dwhen O_2 is the receptor molecule.

This first class is divided in subclasses based on the donor group that undergoes oxidation. For example: 1 denotes a -CHOH group, 2 a -CHO and so on. The third number of the E.C code corresponds to the type of acceptor molecule ($NAD(P)^+$, cytochrome, molecular oxygen, between many others).

E.C. 2: Transferases

Transferases catalyze group transfer reactions. The transfer occurs from one molecule (donor) to another molecule (acceptor). Most of the times, the transfer is performed by a cofactor, that has a transferable group, acting as a donor.

The systematic names of these enzymes are formed according to the scheme *donor:acceptor grouptransferase*, and the common names are given as *acceptor grouptransferases* or *donor grouptranferases.* This class is subdivided according the group that is transferred: a carbon group, an aldehydic or ketonic group, an acyl group, or other one.

E.C. 3: Hydrolases

Hydrolases catalyze reactions involving the hydrolysis of the substrate. These enzymes catalyze the break of C-O, C-N, C-C bonds, as well as other type of bonds.

The systematic name of this enzymes always includes *hydrolase* and the common name is always formed by the name of the substrate with the suffix *ase.* This class of enzymes catalyzes not only the hydrolytic removal of a particular group from a molecule, but it also transfers this group to the suitable acceptor molecule. Taking this in account all hydrolases can be classified as transferases, because hydrolysis can be considered as a transfer of a specific group, being the water the acceptor. Nevertheless, this specific type of enzymes is classified in a different class. In this class, the subclasses are divided by the nature of the hydrolyzed bond (esterases, glycosidases, proteases, and so on).

E.C. 4: Lyases

Lyases are enzymes that catalyze reactions in which functional groups are added to double bonds or, in the reverse reaction, double bonds are formed by the removal of functional groups.

Their systematic name is constructed as *substrate group-lyase*. In cases where the reverse reaction is more important, the enzymes are denominated as *synthases*. The subclasses of these enzymes are divided by the type of bond that is broken during the reaction (Carbon-carbon lyases, carbon-oxygen lyases, between others). The third number of the E.C. code is defined taking in account the elimination group.

E.C 5: Isomerases

This class of enzymes catalyzes the transference of functional groups within a molecule producing isomeric forms, allowing for structural and geometric changes of a compound. In some enzymes of this group, the internal conversion corresponds to an oxidoreduction, however, in these cases, the same molecule acts as hydrogen acceptor and donor. The non-existence of an oxidized product is the reason that justify the falls of these enzymes under this classification, and not as oxidoreductases.

The subclasses of these enzymes are formed according to the type of isomerism, and the next subdivision is performed according to the substrate type.

E.C. 6: Ligases

Ligases are enzymes that catalyze the joining of two molecules with the hydrolysis of a diphosphate bond in ATP, or another similar triphosphate molecule.

The systematic name of these enzymes are *X:Y ligase*, being X and Y the substrates. In this class of enzymes, the second number of the E.C code is related with the type of bond formed (C-O, C-S, C-N bonds and so on).

Several enzymes operate in more than one reaction (promiscuous enzymes) and in these cases, the E.C. number is difficult to define. Nevertheless, the E.C. classification is now a common routine to divide and classify enzymes

Table 1.1 International classification of enzymes: Six types of enzymes according to reaction type

Enzyme Class	Type of reaction
E.C. 1 Oxidoreductases	$A_{red} + B_{ox} \rightleftharpoons A_{ox} + B_{red}$
E.C. 2 Transferases	$A - B + C \rightarrow A + B - C$
E.C. 3 Hydrolases	$A - B + H_2O \rightarrow A - H + B - OH$
E.C. 4 Lyases	$A - B \rightleftharpoons A + B$
E.C. 5 Isomerases	$A - B - C \rightleftharpoons A - C - B$
E.C. 6 Ligases	$A + B + ATP \rightleftharpoons A - B + ADP + P_i$

1.3 A perspective on enzyme catalysis

Having in mind the title of this thesis *"The origin of the catalytic power of enzymes"*, and regarding the diversity of the reaction that enzymes catalyze, one can question: how these different enzymes are able to lower the energy of the biological reactions? Have all these classes similar strategies to decrease the barrier of the biological reaction, or are they different? We will explore these questions in **Chapter 3** (*Activation free energy, substrate binding free energy and enzyme efficiency fall in a very narrow range of values for all enzymes*) with a detailed analysis of some parameters related with the catalytic power of enzymes, in each enzyme class presented before.

Careful measurements of reaction rates in water, and its comparison with the rates of the same reaction catalyzed by enzymes, have been showed that these biocatalysts enhanced the rates of the reactions by 10^{15} to 10^{17} fold. These values, which were presented and well discussed by Wolfender [1,12], provide a measure of what it meant by a catalytic process and justify why enzymes are considered the best catalyst ever.

The environmental differences between a transition state in an aqueous solvent shell or surrounded by the amino acid residues on the active site, are very large. The catalytic effect promoted by enzymes will always be originated by a greater transition state stabilization (low activation energy), compared to the reference reaction in water. The big question is what kind of forces are responsible for this catalytic power. Are they electrostatic, steric, promoted by hydrogen bonding, by (de)solvation effects, by entropic effects, or by many of these effects at same time?

Many studies defend that the catalytic residues are precisely positioned in the active site to allow the chemical reaction to occur [13,14]. However, how this alignment is achieved? Other authors extrapolate that a preorganization of the active site allows the selection of

substrate subpopulations, which approach the configuration of the transition state and bound the active site with higher affinity. If it is the most important argument to justify the catalytic power of enzymes, the energy associated with substrate binding should be the most important parameter in bio-catalysis. Relating to this theory, some authors suggests the existence of near attack conformations (NACs), that corresponds to those "subpopulations" of substrate conformations that favor the catalysis and whose ligand distances/angles are defined to facilitate the reaction [15-17].

Other studies suggest that instead of steric/conformational effects, the catalysis is controlled by electrostatic effects [18-20]. A more general perspective supports that a combination of steric and electrostatic interactions works together to align the enzyme active site, substrates and cofactors to promote catalysis. Other perspectives defend that the large number of degrees of freedom in macromolecules, makes them very flexible, and these motions may aid catalysis [13]. All these theories were briefly explored further in this thesis.

1.3.1 Transition State theory

Having in account the hypothesis proposed to explain the origin of the catalytic power of enzymes, many theoretical and experimental studies have been performed to try to corroborate the suggested ideas.

In this thesis, different catalytic mechanism of enzymes will be presented. All these processes are characterized by reaction rates, that defines how fast a reaction takes place. A schematic representation of an uncatalyzed and an enzyme catalyzed reaction is show by **Figure 1.1** indicating some of the important quantities used in the characterization of reactions.

According to the transition state theory (TST) [13,21,22], a chemical reaction proceeds by a continuous change, in the relative positions and potential energy, of the constituent atoms. On the reaction path, between the reactants and products there exists an intermediate configuration at which the potential energy has a maximum value. This structure is known as an activated complex and its state is referred as a transition state (TS), which corresponds to a saddle point in the energy surface. For uncatalyzed reactions, the activation barrier corresponds to the difference between the transition state energy and the initial structures (reactants), which corresponds to a relative free energy ($\Delta G^{\ddagger}$). In the enzyme-catalyzed reactions, before the chemical reaction, the substrate S, binds the active site of the enzyme E, forming a ES complex, which is energetically more stable. Then, the reaction proceeds to the chemical step, to form the product, which leaves the enzyme in the end of the reaction (P+E). The activation barrier, in this case, is calculated

relatively to the complex ES. The reaction rates are usually determined experimentally, while computational calculations are largely used to describe the potential energy surface and the relative energies between different states.

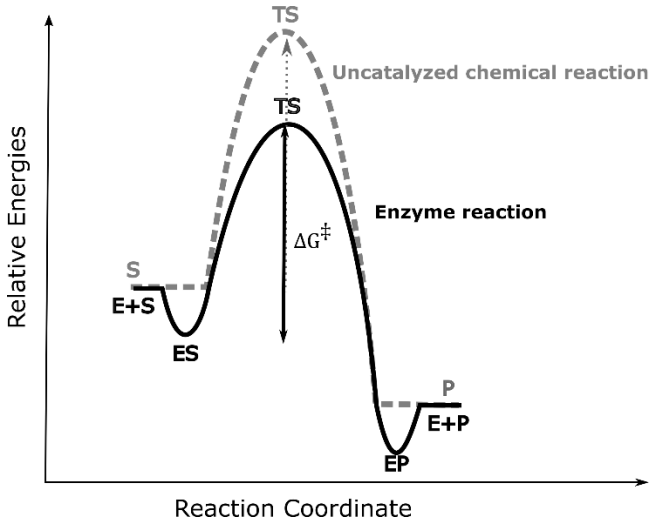


Figure 1.1 Schematic representation of the free energy profile for uncatalyzed (grey) and enzyme catalyzed (black) elementary reactions.

TST assumes that a reaction does not go back again, after it passes through its activation barrier. It assumes that there is an equilibrium energy distribution at all states (stable and unstable) along the reaction coordinates. As predicted by the Eyring equation, the rate of the product formation (rate constant), k, depends on the reaction temperature, and can be given by the following equation:

$$k = \frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger}}{RT}} \qquad \textbf{Equation 1.1}$$

In this equation, T represents the temperature, k_B is the Boltzmann's constant, h is the Planck constant and R is the ideal gas constant. If the rate constant of a reaction is known, TST can be used to estimate its standard Gibbs energy of activation.

The term $\frac{k_B T}{h}$ indicates the number of collisions between reacting particles at a given temperature, and $e^{-\frac{\Delta G^{\ddagger}}{RT}}$ predicts the number of particles with enough kinetic energy to overcome the transition state. This last term was derived for ideal gases, though it is usually applied also for condensed matter (such as enzymes), and it is a consequence of the distribution of kinetic energy as a function of temperature (Maxwell-Boltzmann distribution).

Looking to the equation in both terms, temperature always increase the rate of the reaction, increasing the number of collisions in the first term, or increasing the number of particles with sufficient kinetic energy to overcome the barrier. In other hand, if the activation entropy was too negative, which means that there a small number of states associated with the transition state, an increase in the temperature will increase the Gibbs free energy barrier, taking in account the following equation:

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger} \qquad \textbf{Equation 1.2}$$

In case of a multistep mechanism, the reaction rate is defined by the transition state with the highest activation barrier, which corresponds to the rate-limiting step.

1.3.2 Linking theoretical calculation to kinetic experiments

Computationally, enzymatic catalysis is one of the most complicated processes to study and simulate, because of the large size of the models, the timescales and the levels of theory involved. As seen above, to study all the process associated with enzyme catalysis it is necessary to take into account, not only the chemical step, but also the initial binding of the substrate to the enzyme, and the final release of the product.
Enzyme kinetics study the rate of a reaction catalyzed by enzymes, as well the parameters that affect this rate. Kinetics experimental studies can be analyzed taking in account the Michaelis-Menten formalism, using the following scheme, to model the enzyme action:

$$E + S \underset{k^{-1}}{\overset{k_1}{\leftrightarrow}} ES \xrightarrow{k_{cat}} EP \rightarrow E + P \qquad \textbf{Equation 1.3}$$

Where $\frac{k^{-1}}{k_1}$ represents the affinity of the enzyme *(E)* for the substrate *(S),* to form the complex enzyme:substrate (*ES*), and k_{cat} represents the rate constant for the transformation of the substrate into the product (*P*). All these steps occur in different timescales. For example, the bond forming and breaking occurs in a time scale compared to the molecular vibrations, being the life time of a transition state in the order of picoseconds, or less. In these cases, the system needs to be treated with an electronic method (See **Chapter 2**) to describe the electronic occurrences. In turn, to study the substrate binding, or the product release, that are processes that, in principle, take more time (microseconds to milliseconds), molecular mechanics (MM) methods are reasonable, however a large conformational space exploration is needed. An inappropriate sampling due to a non-equilibrated MD, or insufficient time of simulation, or the use of a wrong force field, could lead us to make mistakes during our computational calculations.

The comparison between theoretical and experimental data are very important to validate our models. With experimental methods the rate at which the reactants are converted into the products could be defined as:

$$v = \frac{d[P]}{dt} \qquad \textbf{Equation 1.4}$$

This kinetic information is measured on condition of enzyme saturation ($[S] \gg [E]$) , with all the molecules in form ES. The rate obtained in those conditions is denoted as $v_{máx}$ (the maximum velocity of the enzyme). From kinetic studies, the previous mentioned k_{cat} , or turnover number, is also calculated, corresponding to the number of substrate molecules, turned over per enzyme molecule, per unit of time (s^{-1}). For example, a k_{cat} of 15 s^{-1} means that an enzyme catalyzes in average 15 reactions each second.
K_M is another important value provided by kinetic studies. This parameter is related with the enzyme affinity to substrate. This value corresponds to the concentration of the substrate at which the reaction rate is half of $v_{máx}$. Enzymes with a high value of K_M, required a higher concentration of the substrate to reach a given reaction velocity. The ratio between k_{cat} and K_M is referred as a “specific constant” and it is usually used to measure the overall efficiency of the enzyme. All these values could be related in the following expressions:

$$v = \frac{d[P]}{dt} = v_{máx}\frac{[S]}{K_M + [S]} = k_{cat}[E]\frac{[S]}{K_M + [S]} \qquad \textbf{Equation 1.5}$$

[P], [E] and [S] are the concentration of the product, the enzyme and the substrate, respectively. The kinetic parameters k_{cat} and K_M , which were described above, are generally useful to study and compare different enzymes and they are very informative for computational enzymatic studies. These parameters also allow the evaluation of the kinetic efficiency of enzymes.

1.4 The origin of the catalytic power of enzymes

As seen until now, kinetic results could help to understand the rates of the enzyme reactions, however, how these molecules accelerate difficult chemical transformations, that are, most of the times, impossible in water? Some enzymes evolved by optimizing the value of k_{cat}/K_M [1], however the main question stills related with the decrease of the activation barrier ($\Delta G^{\ddagger}$). It is not clear, until now, how different factors (such as electrostatic

effects, steric effects, hydrogen bonding, solvation effects, between other pointed above) are able to contribute to the big catalytic power of enzymes.

In this section we will analyze the current level of understanding and some proposed hypothesis to answer this question. The choice of presented hypothesis reflects our own perspective, and, because of that, other interpretations and some other hypothesis could be explored and considered to explain the origin of the catalytic power of enzymes.

Until now, none experimental, neither theoretical study, explain the origin of the catalytic power applied to all enzymatic reaction, in general, but rather, they try to present the best explanation for particular systems.

There are different factors that are described as possible sources of catalysis [15,17,19,23-27] and some of them will be described in the next lines.

1.4.1 Desolvation hypothesis

In solution, for a chemical reaction to occur, the molecules of the bulk solvent must reorganize their positions around the reactive species, to facilitate a configuration that will allow the formation of a transition state and, consequently, the progress of the reaction to the products. Some authors suggest that these reactions, in solution, are retarded by the solvent reorganization, since a great amount of energy is spent in the reorientation of the solvent molecules [28] .

Many times, enzymes active sites are buried within the protein, protected from the polar water solvent. Other times, when the substrate binds the enzyme active site, the solvation is weakened, and the water molecules are forced to go out, resulting in an environment without (or with less) water molecules, which seems to accelerate the reaction. This hypothesis suggests that the polar nature of the active sites seems to interact with the substrate, without the necessity to reorganize itself between many positions and, because of that, without the great cost associated to this necessity in solution [29,30]. This is also corroborated by the low dielectric environment typical of an active site.

Some authors also propose that enzymes work by providing a non-polar, or gas-phase like, environment that destabilizes highly charged ground states [31,32]. However, this hypothesis is less acceptable, because, as pointed by Warshel [19], a polar TS is less stable in non-polar active sites, than in water.

1.4.2 Electrostatic effects

Warshel and his collaborators defend, in turn, that electrostatic effects are the main responsible for enzyme efficiency as catalysts [18,33].

It is well known that the environment has a big influence on the progress of a chemical reaction. The dielectric constant and the polarity of the medium are factors that largely influence the reaction rates.

Enzyme's active site are usually heterogenous and very polar, been able to create very intense and oriented electric fields. Because of that, some properties of the substrate can be affected by the residues present on the active sites, such the pK_a, for example. Furthermore, as mentioned above, in solution, the energetic cost associated to the reorganization of the solvent molecules, to orient them to promote the reaction, is very high, while the electrostatic environment is already well organized in enzyme's active site. Many authors defend this preorganization of the active site residues, which has well oriented dipoles, facilitate the orientation of the reactants to the transition states, without any reorganization energetic cost [34-38] .

However, though this hypothesis seems plausible, the catalytic power of enzymes should result from the combined contributions of several factors, instead of a single one.

1.4.3 Entropic effects

Other authors suggest that enzyme catalytic power can be justified by a less entropic cost, when compared to the reactions in aqueous solution [39-41]. In solution, a large energetic cost is associated with bringing two reactants together, which leads to a large penalty in the entropy of the system. The enzyme active site preorganization drastically reduces the configurational space available for the positioning of the substrate, and consequently, it reduces the energy needed to maintain the reacting molecules in a good orientation to react. Although this alignment of the reacting molecules is entropically not favorable, the enzyme restricts the conformations to explore to achieve the transition state and continue the reaction.

The entropic loss can be compensated by the enthalpy of binding and by the favorable energetic contribution of the release of the water molecules to the bulk solvent. This entropic effect was reported as the responsible for the catalytic effect of some enzymes [42,43], however, it was also showed, by other studies, that this effect is very small, or even non-existent for other enzymes [1].

1.4.4 Destabilization of the ground state

In some papers, the origin of the catalytic power of enzymes is associated with a destabilization of the ground state, promoted by a specific special disposition of the active site residues that imposes physical constraints on the substrate [44-47].
Ground state destabilization is a thermodynamic concept, which suggests that to go from one state to another, there is an increase in energy. Applied to enzyme catalysis, this concept could explain the efficiency of these molecules, by a destabilization of the substrate, when it binds to the active site. This destabilization promotes a reduction of the activation energy, increasing the reaction rate.
We present the ground state destabilization as an isolated hypothesis because many published works refer to it to explain the origin of the catalytic power of enzymes, however this hypothesis can be related with the previous ones. For example, the ground state destabilization can occur, for example, by an electrostatic effect promoted by the active site preorganization.

1.4.5 Orbital steering

Other suggestion to explain the increase of the reaction rates by enzymes is the “orbital steering”. Koshland and co-workers [48,49] postulated a very strong dependence of reaction rate on the precise orientation of atomic orbitals on the reacting atoms (orbital steering). This hypothesis suggests that the catalytic activity of enzymes depends on the ability, not only to approximate the reacting atoms, but also to steer their orbitals throughout a path which takes advantage of this direction preference, or, in other words, approximate reacting species in the most favorable orientation.

1.4.6 Low barrier hydrogen bonds

Hydrogen bonds have an essential role in protein structure. Common hydrogen bonds involve -NH or -OH groups as donors and atoms of N or O as acceptors. The length of these hydrogen bonds in water, is near 2.8 Å and they energy of formation is near 5 kcal/mol. The concept of low barrier hydrogen bonds (LBH) arrived due to the existence of other type of hydrogens bonds, usually shorter (with distances of around 2.5Å) and having energies of up to 20 kcal/mol. The formation of this short and very strong LBH in the transition states or in the intermediate complexes, seems to be an important contribution to enzyme catalysis. It is interesting to note that this type of bonds is not common in water, however, the confined environment of the active site, allows this type

of LBH bonds. An enzyme is able to convert a weak hydrogen bond into a strong one, changing the *pK* value of the substrate, which becomes closer to that of the enzymatic group to which it is hydrogen bonded [50-52].

In turn, Warshel and co-workers, affirmed that LBH bounds prevent some reactions instead of promoting them, because their distribution of charge is more diffuse, resulting in a lower stabilization of the transition state [53,54].

1.4.7 Preorganization of the active site and Near attack conformations (NACs)

A preorganization of the active site residues allows the selection of specific populations of substrates that will bond in a very specific position with high affinity.

After binding, the substrate should adopt a ground state conformation that facilitates the progress to a transition state structure. These conformations are known as near attack conformations (NACs). The preorganization of the active site residues will favor in a large way the formation of NACs, compared with those in solution, which could be the key point to justify enzyme catalysis [17]. This hypothesis is, in part, dependent on dynamic fluctuations of the enzyme-substrate complex, changing the interactions between the substrate and the enzyme, resulting in a correct alignment to favor the progress of the reaction to the transition state. Some studies showed that the occurrence NACs is larger within the enzyme active sites than their presence in bulk water. For example, in the reaction catalyzed by the enzyme chorismate mutase, it has been showed that the free energy associated to the formation of a NAC in water is near 8 kcal/mol, while it decreases to 0.6 kcal/mol in the enzyme [55,56]. This result shows evidence for a thermodynamic stabilization of NACs over the ground state, compared with the uncatalyzed reaction. There are other studies that also reports the formation of NACs on the same enzyme, but with only moderate contributions to its catalytic power [57,58].

1.4.9 Correlated structural fluctuations

Between all the several proposals that have been suggested, the idea that enzyme dynamics somehow contributes to catalysis has been largely discussed during the recent years [25,59-64]. It is important to refer that the concept of enzyme dynamics is treated in different ways by different authors.

The logical relationship between protein motion and function is a fundamental question for both biology and physical chemistry. Atoms are not static, and they move during any

chemical process that occurs at above few ºK. Taking this in account, we cannot relate catalysis with the simplest fact that atoms are moving. However, it is well-supported that protein structures exhibit dynamical fluctuations on a wide range of timescales. These timescales start from femtoseconds (10-100 fs) characteristic from bond vibrations, passing by picoseconds (10-100 ps) for rotations of side chains at the protein surface, 1 - 10 ns timescale characteristic from the rotation of residues at the interfaces, increasing to micro and millisecond timescale for rotation of medium sized side chains in the protein interior and large domain motions [65]. The fundamental question, that will be explored in some works addressed in this thesis, is that if there is a relation between these dynamic fluctuations of the enzyme structure and their catalytic activity and, if so, what is the nature of this relation.

Karplus and McCammon stated one of the clearest explanations of this possibility. They postulated that "*If the substrate is relatively tightly bond, local fluctuations in enzyme could couple to the substrate in such way to significantly reduce the barriers. If such couple effect exist, specific structures could have developed through evolutionary pressure to introduce directionality and enhance the fluctuations…".* They also affirm that "*Energy release locally in substrate binding may be utilized directly for catalyzing its reaction, perhaps by introducing certain fluctuations*" [66].

It is well accepted that enzyme dynamics can influence enzyme activity in several ways [67]:

- influencing the *structural stability* of an enzyme, favors an active conformation rather than others;

- influencing the *binding affinity* between an enzyme and its substrate;

- modulating the fine structure and the thermal stability of the transition state of the complex enzyme-substrate, altering, consequently, the activation energy of the catalyzed reactions.

Taking this in account, protein dynamics seems to have influence on both enzymatic kinetics and protein conformational space.

There are many studies in different enzymes that support the linkage between protein structural fluctuations and catalysis. Between them, the case of Catechol O-Methyltransferase, that catalyzes a methyl transference from S-adenosylmethionine to the hydroxylate oxygen of a substituted catechol. The authors defend that the catalysis promoted by this enzyme is associated with a general movement of the protein, as a

consequence of the approximation of a magnesium cation to the nucleophilic group [68]. In the case of chalcone isomerase, for example, the authors found significant differences between the reaction catalyzed by this enzyme and the equivalent reaction in aqueous solution. Intramolecular changes were observed inside the enzyme active site, and those dynamic motions were considered as important to the evolution of the reaction along the reaction coordinate [69]. Mulholland and some co-workers, discussed the role of dynamic in DNA polymerase, showing that some authors defend that there is a correlation between some pre-chemical conformational changes of this enzyme and its catalysis, while other authors do not assign any influence of these conformational fluctuations on the control of the free energy of this reaction. These are good examples of the ongoing debate on the potential role of protein motions on catalysis.

Some other studies discussed the role of conformational fluctuations [23,70,71], however the two most discussed cases are the dihydrofolate reductase (DHFR) [72-76] and liver alcohol dehydrogenase (LADH) [77,78].

In the case of DHFR, the first evidences of the structural flexibility derived from crystallographic studies. X-ray crystal structures of DHFR, reveal that this enzyme can assume different conformations (open, closed or occluded), depending on the nature of the bounded ligand. A conversion between a closed to an occluded state was observed, during the reaction. This fluctuation occurs due to the presence of a loop, that changes its position to facilitates the chemical reaction, which supports the influence of structural motion in catalysis. In this enzyme, these conformational transitions require large-scale reorganization of the different regions of the protein. NRM studies showed large-amplitude picoseconds to nanoseconds timescale motions that are reduced after the closure of the loop. Experiments on a large timescale (microseconds to milliseconds) showed also that there are motions, at these time scales, directly linked to catalysis. Hybrid quantum mechanics/molecular mechanics calculations and classical molecular simulations corroborated the experimental studies, showing that some fluctuations occur on the femtosecond to picosecond timescale. These studies illustrate the importance of structural fluctuations and conformational sampling to study enzyme catalysis.

In a similar way, LADH also exhibits conformational changes that facilitate its catalytic mechanism. Experimental and theoretical studies showed that, the catalytic domain rotates upon the formation of enzyme-substrate complex, and a flexible loop rearranges its structure to accommodate the changes from an open to a close form, similar to DHFR. All these works show that fast motions (in picosecond to nanosecond timescale) have restraints imposed by the protein fold and create configurations of the protein-substrate complex, that facilitate the chemical reaction to occur. Sometimes, these fast motions seem to be in equilibrium as the reaction processes along the reaction coordinate. Larger

order motions seem also to influence the catalysis, however there is no concrete evidence of the truth of this type of coupling.

Despite the previous examples, Warshel and co-workers stated that the reactive fluctuations are similar in enzyme and in solution, defending that no dynamical effect has ever been experimentally shown to contribute to catalysis [19,25,79-81].

Regardless of the contrary opinion of different authors, this hypothesis has been debated during the last years. In this thesis some emphasis will be given to these ideas during the **Chapter 7** and **8**.

1.4.8 Quantum effects

A chemical reaction implies the conversion of reactants to products, throughout a transition state, and with an associated barrier ($\Delta G^{\ddagger}$). When the quantum effects are considered, the probability that the system can tunnel through the barrier without any energetic [81] cost is different from zero. Tunneling was reported for electron transfer, but also for hydrogen atoms, and this later effect has been suggested, by different authors, to play a role in enzyme catalysis [82-84] .

This effect is strictly related with the previous presented hypothesis of the dynamics in catalysis, because the organization of the active site and their motions could promote the well alignment between the energy of the reactants and products in order to promote the hydrogen tunneling. For example, for DHFR, quantum effects were also reported, associated with specific motions of this enzyme [85-87].

In summary all the strategies presented here to justify the catalytic power of enzymes have been defended by some authors and rejected by others. In fact, some of these theories are related between each other, and some of them seem to have more theoretical and experimental support than others, such as electrostatic stabilization, with a pre-organization of the active site, and structural fluctuations (dynamic effects). However, until now, there is no clear explanation about the origin of the rate enhancement of the reactions catalyzed by enzymes.

1.8 The main goals of this thesis

Enzymes are essential to life, processing their substrates with astonishing kinetics. Understand how enzymes do this is one of the major question of biochemistry. As pointed

above, the precise origin of this large catalytic power is still unknown. This topic has been largely debated and the results of different authors, in different enzymes, lead to different conclusions, all undemonstrated and not consensual, as we pointed above.

The goal of this thesis was studying the origin of the catalytic power of enzymes. The next chapters present different works, that could contribute to the fundamental understanding of enzymes, their catalytic mechanisms and, ultimately, to give some tasks on the understanding of biocatalysis.

One of the main problems related with this field is that the large part of the hypothesis, presented in the previous subchapter, were derived from the study of few specific enzymes. Sometimes, many years are needed to understand how a single enzyme works. Consequently, when comes the time to look to the origin of the catalytic power, the authors concentrate their studies in enzymes for which they have knowledge and know-how. However, enzymes are extremely diverse, and a conclusion based on a few specific enzymes cannot be valid on a global scale.

Having this in mind, one of the first tasks of this thesis was to build and analyze a "general database" with rates and Michaelis constants for a large number of enzymes, as well as with structural and mechanistic information about them (such as their size, oligomerization state, presence of cofactors, type of reactions that they catalyze, promiscuity, etc.). The main objective of this database would be to help to correlate all these collected properties, with the rate constants and Michaelis constants, to identify the patterns responsible for enzyme efficiency. The results of this work are presented in **Chapter 3**.

Another objective of this work was to describe the catalytic mechanism of different enzymes using QM/MM calculations. In **Chapter 4** and **Chapter 5**, the catalytic mechanism of Human Renin and Macrocyclase PatGmac are described with atomistic detail.

In parallel we also evaluated the influence of some protocol questions that could influence the results obtained by computational methods that we used. In **Chapter 6** we present the influence of frozen some residues, during a QM/MM study.

Between the initial objectives of this thesis was also to study the potential energy surface for the rate-limiting step of different enzymes (with a well-known catalytic mechanism), starting from different initial structures, to account for the influence of enzyme flexibility on catalysis. In **Chapter 7** and **Chapter 8** we perform this kind of analysis for α-amylase and for HIV-1 protease. Through all these works, we hope to contribute to a better understanding of enzymes and give also important keys to understand their catalytic power.

The next chapter will summarize the theoretical background used to perform all these studies. The remaining chapters present the different works studied during these four

years. They are not shown by chronological order, but by the order that we found more logical to connect all of them.

CHAPTER 2. Computational methods to study enzyme catalysis

2.1 Introduction

Understanding how enzymes work at the molecular level is a fundamental problem in biochemistry. These biomolecules are very efficient and specific, but despite intensive experimental and theoretical works on this field, the origin of their rate accelerations remains unclear, as pointed in the previous chapter.

Study enzyme catalysis is not only important for fundamental and theoretical biochemistry, but also for the development of new efficient catalysts, understanding the activity of enzyme mutants, as well as for the rational design of new proteins and inhibitors.

As it will be demonstrated by the works that will be presented in this thesis, computational simulations are a good tool to study enzymes and their reactions. These methods can provide information which is often inaccessible by experimental works. The increase in computational power during the last years associating with the development of more efficient algorithms, allows the study of large molecules, as enzymes [88,89]. Though, computational chemistry/biochemistry needs always a good compromise between the accuracy of the methods, the size of the models and the time that is needed to describe the studied phenomena.

Enzyme catalysis involves, at least three different stages:

1) Substrate binding;

2) Chemical Reaction (performed in one or more chemical steps);

3) Product release.

The large part of this thesis is focused on the chemical reaction steps and the methods that will be presented are, manly, for the study of this stage. We will not discuss methods used to study the substrate binding, neither the product release.
The chemical reaction steps involve bond breaking/making, that occur in a very fast time scale, being that the transition state structure has a half-life in the sub-picosecond time-scale. Structural and energetic details of these unstable species are often not experimentally detected.
In turn, study enzymatic reactions using computational methods is still challenging. The large size of these molecules, the timescales of the reactions that they catalyze (femtoseconds to nanoseconds order) and the levels of theory involved, are obstacles to overcome. Specialized computational methods are required, and it is possible to divide them in two main levels of theory:

- Classical Molecular Mechanics (MM);

- Quantum Mechanics (QM).

From now, classical molecular mechanics (MM) methods are the implemented method to study systems throughout nanoseconds to microsecond-timescale. The use of MM methods alone cannot be applied to describe catalysis, due to the neglection of explicit electronic rearrangements, however it can be useful to perform an extensive conformational sampling of enzymes and give important thermodynamic and structural insights about the system [90-93].
To study enzyme catalysis the use of QM methods or hybrid QM/MM methods are the most efficient approaches. QM/MM methods allow the decreasing of the computational costs, associated with QM calculations, without compromise the accuracy of the chemical description of the system [88,94-96].
With this view in mind, in the next subchapters we will briefly described the computational methods used to study enzyme catalysis, their advantages/pitfalls and applications. Our perspective will give more emphasis to QM/MM methods, because they were largely used in the works that will be present in the next chapters.

2.2 Molecular Mechanics

As mentioned above, biochemical molecules are too large to be studied with QM methods, which are time-consuming due to the explicit description of the electrons of the system. In contrast, MM methods (also known as force field methods) are defined by simple

equations based on classical mechanics, which describes the molecules as an aggregate of spherical particles coupled by harmonic springs. In these methods, the atoms are treated as the smallest component of the system and the protons and electrons are not described explicitly. Parameters such as atomic charges and van der Walls radius are used to approximate their effects. To calculate the energy of the system, the potential that describes the interaction between particles is parceled into different terms using simple harmonic/sinusoidal potentials [91,92,97].

The main advantage of the MM methods is, in fact, the possibility to describe systems with a significant number of atoms, without consuming a large amount of computational time. This is possible due to the simplicity and parameterization of the potential energy function, which becomes fast to solve, even for a large number of atoms. These are the methods of choice for protein simulation in which the conformational flexibility is being studied. In turn, MM methods cannot provide properties that depend on the electronic distribution, as breaking or making bonds

2.2.1 Force Fields

This subsection will focus on the most generic biomolecular empirical force field, that describes the potential energy function (*V*) as a sum of bonded and nonbonded energy terms, divided in six terms, as given by Equation 2.1 and Equation 2.2. The AMBER force field [98] which was employed in the works that we present here, uses these equations.

The bonded terms can be written as:

$$E_{bonded} = \sum_{bonds} \frac{1}{2} k_l (l - l_0)^2 + \sum_{angle} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{\substack{improper \\ torsions}} \frac{1}{2} k_\varphi (\varphi - \varphi_0)^2 + \sum_{dihedrals} K_\phi [1 + cos(n\phi - \gamma)] \quad \text{Equation 2.1}$$

The nonbonded terms can be represented by:

$$E_{nonbonded} = \sum_{nonbonded} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + \sum_{\substack{nonbonded \\ pairsij}} \varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] \quad \text{Equation 2.2}$$

2.2.1.1 Bonded Terms

The force field comprises four types of bonded interaction: bond stretching terms, angle bending terms, dihedral or torsional terms and improper dihedrals.

Bond stretching describes the type of bond between two atoms (if it is simple, double, triple or aromatic). To simplify, force field approximate this interaction to a harmonic potential described by two parameters: the equilibrium bound length l_0, and the force constant k_l, that describes the harmonic regime near the equilibrium, as can be seen by Equation 2.3.

$$E_{bound} = \frac{1}{2} k_l (l - l_0)^2 \qquad \textbf{Equation 2.3}$$

These terms are variables in the quadratic potential energy equation for bonds E_{bound}, being l the distance between two atoms. The quadratic approximation is appropriate for conformations near the minimum. When the distance decreases, the potential energy increases very fast, due to the Pauli repulsion between the core of electrons of two near atoms. The E_{bound} energy of the stretching between two atoms should increase quickly and then converge slowly to the point when there is no interaction.

The potential energy associated with the *angle bending* between three bonded atoms is also described by a quadratic equation, as represented by Equation 2.4.

$$E_{angle} = \frac{1}{2} k_\theta (\theta - \theta_0)^2 \qquad \textbf{Equation 2.4}$$

The k_θ represents the force constant, θ is the angle and θ_0 corresponds to the equilibrium angle. This harmonic potential is a good approximation over a large range of angles, however, it may give more incorrect results for some types of angles. For example, the appropriate chemical behavior for linear bond angles will not be correctly described by this potential. Another defect of Equation 2.4 is that for some cases (some inorganic systems, for example) it is possible the existence of multiple equilibrium values.

The next two terms of the equation are *dihedrals and improper torsions*. These terms are more difficult to parameterize since, in molecules, there are many possible dihedral interactions between groups of four atoms linked by harmonic springs. These interactions are classically treated as rigid rotors in which the 1st and 4th atoms rotate around the bond formed by 2nd and 3rd atoms. These parameters exhibit a periodic behavior every time a complete rotation occurs. The most accurate way to treat this parameter is to consider a Fourier expansion of a sinusoidal potential, where each minimum and maximum are well defined. Taking into account the computational cost and the small values of the energy

associated with these transitions, only the first term of the Fourier expansion is considered in the Amber force field, as represented by Equation 2.5:

$$E_{dihedrals} = \sum_{dihedrals} K_{\phi}[1 + cos(n\phi - \gamma)] \quad \textbf{Equation 2.5}$$

Three different parameters are necessary to calculate this energy: $K_{\phi,}$ represents the maximum energy that describes the torsion, n is the number of minima of a complete torsion, ϕ represents the current dihedral angle, and γ corresponds to the phase with lowest potential energy.

Improper torsions also involve groups of 4 linked atoms. In this case torsions resulted from strain of *π*-molecular orbitals that are formed from the binding *sp*^{2}-hybridized atoms, which are in a planar conformation and cannot be described by the potential energy described above. As *bond stretching* and *angle bending energy*, the energy associated with improper torsions are described by a harmonic potential, as represented by Equation 2.6. The coordinate that is approached is the angle that is formed by two non-concurrent vectors in the plan (φ).

$$E_{\substack{improper \\ torsions}} = \sum_{\substack{improper \\ torsions}} \frac{1}{2} k_{\varphi}(\varphi - \varphi_0)^2 \quad \textbf{Equation 2.6}$$

2.2.1.2 Non-bonded terms

As referred above, electrons are not described explicitly in a molecular mechanics force field. Instead of that, they are considered in an approximately way by three parameters: atomic charge q, minimum Lennard-Jones interaction energy ε and σ (which corresponds to the interatomic separation at which the repulsive and repulsive energy terms cancel out), that address the role of electrostatic and long-range interactions.

The electrostatic interaction between two non-bonded atoms with partial charges q_i and q_j is commonly described by the Coulomb's potential, as represented by Equation 2.7:

$$E_{ee} = \sum_{nonbonded} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad \textbf{Equation 2.7}$$

where ε_0 is the medium permittivity, and r_{ij} is the distance between atoms i and j. Usually, non-bonded interactions are excluded for 1-2 (two atoms) and 1-3 (atoms that are bonded to a common atom) to prevent numerical instabilities that could result from the short distance interactions, and because the interactions between atoms in 1-2 and 1-3 relative

positions are already taken into account through the E_{bound} and E_{angle}. 1-4 interactions are considered but are scaled by a factor of 0.5 (AMBER force field) or they could be also excluded, depending on the force field.

E_{vdW} (van der Waals energy) describes the long-range interactions and could be interpreted as the non-polar part of the interaction, not being related with the energy provided by the partial atomic charges. This parameter describes two important behaviors of electrons and protons: Pauli repulsion and London dispersion. From the Pauli repulsions it is known that no more than two electrons can occupy the same physical state at the same time. This means that when the atoms become very close, the repulsion between them increase, preventing additional approximation. Dispersive interaction corresponds to the dispersive force between instantaneous dipoles and induced dipoles. These forces are attractive and decay very rapidly with the distance. A 12-6 potential is used to describe vdW interactions between atoms i and j, at a distance r_{ij}, as represented in Equation 2.8.

$$E_{vdW} = \sum_{\substack{nonbonded \\ pairsij}} \varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] \qquad \textbf{Equation 2.8}$$

The term ε_{ij} corresponds to the minimum potential energy and $R_{min,ij}$ is the distance when the energy reaches the minimum (this distance is related to σ as $R_{min,ij} = 2^{\frac{1}{6}}\sigma$). In this equation the positive factor, with a power of 12, represents the Pauli repulsion and the negative factor, with a power of 6, is a model for the London dispersion.

Both the Coulomb's and the Lennard-Jones's potential are slowly converged sums, in particular the former. A truncation criterion should be imposed, to allow a calculation of a finite contribution from the explicit counting of the interactions. In Molecular Dynamics simulation, as we will see in the next section, the periodic boundary conditions (PBC) are combined with a *cutoff* (generally inferior to half of the largest length of the system), to provide a list of atoms that will interact explicitly with each other [1].

[1] The information summarized here was taken from General Books of Computational Chemistry as *Cramer et al.*; *Jensen et al.,* between others.

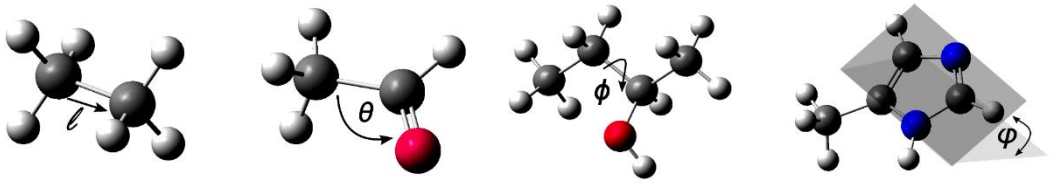


Figure 2.1. Representation of internal coordinates employed in common empirical biomolecular force fields. l represents the bond length, θ corresponds to angle bend, ϕ represents torsion angle and φ improper torsion.

2.2.2 Molecular Dynamics

2.2.2.1 Energy minimization

The first calculation of any molecular modelling work is an energy minimization of the system, to refine the initial structure. During this step, the structure of the studied molecule is modified until an energy minimum is found in the potential energy function described above. In the case of large systems, as biomolecules, the minimization of the system is an essential step taking in account two main aspects:

-First, in biomolecules there are many degrees of freedom and the starting configuration may not be on a stable conformation.

- Second the studied system is usually a result of a modelling process. The X-ray refinement, the subsequent addition of hydrogen atoms, solvent and counter-ions, substrate modelling, could introduce bad contacts (generally non-bonded vdW and electrostatic interactions), that needs to be improved.

The common procedure to minimize the energy of the system is a graduate relaxation. In the systems presented in the next chapters (enzyme-substrate systems with solvent), the most common protocol comprehends a first minimization of the solvent molecules, followed by a minimization of the hydrogen atoms of the system (added by a computational algorithm) and, lastly a minimization of the full system. Sometimes an intermediate step, with a minimization of the protein backbone, precedes the last step. These separated steps help to solve the previous described problems related with bad contacts on the modelled system.

To address the problem of finding an energetic minimum, two algorithms are commonly used: *steepest descent* [99] and/or *conjugate gradient* [100]. The *steepest descent* algorithm is good for rapidly removing the largest strains in the studied system, however it also converged slowly when close to a minimum. In this case the *conjugate gradient* is more

efficient. The *steepest descent* method starts at a point x_0 (initial guess) and moves from x_i to x_{i+1}, as many times as need. This algorithm moves coordinates of the system in the opposite direction from the gradient of the potential energy function. At each iteration, the algorithm moves the coordinates according to [101]:

$$x_{i+1} = x_i - \varepsilon \nabla V(x_i) \qquad \textbf{Equation 2.9}$$

being the x_{i+1} and x_i, the new and the previous position, respectively. ε corresponds to the size of the step taken towards the minimum and $\nabla V(x_i)$ represents the gradient of the potential energy along the x_{i+1} to x_i direction. Due to the slowly convergence of this algorithm, it is usual to change *t* to a *conjugate gradient* algorithm when the system is close to a minimum. In this case, instead of one gradient, a conjugate of gradients is used, using all gradients from the previous steps. The main advantage of this algorithm is a more focused search for a minimum, preventing the possibility of the search vectors to point back and forth, albeit the computing time is bigger than for the *steepest descent* algorithm.

2.2.2.2 MD simulation

After the minimization step, the system reaches a minimum in the potential energy surface. However, the obtained configuration is only one between the large number of configurations that system can adopt. Molecular dynamics (MD) methods produce a trajectory, or in other words, a series of time-correlated microstates, by propagating an initial set of coordinates and velocities according to Newton equations by a series of finite time steps.

MD is deterministic, which means that if one starts two MD simulations with exactly the same initial coordinates and velocities, the obtained trajectory will be the same.

The microstate which an equilibrium MD simulation reproduces can be characterized by state functions such as volume (V), pressure (P), total energy (E), temperature (T), number of particles (N), chemical potential (μ) and others. The generated trajectory corresponds to an ensemble of different microstates for a given set of fixed values of these variables. MD simulation is a way to explore the ensemble of possible microstates along time by solving the potential energy function of the system (presented above in the force field equation) in conjunction with the Newton's laws. In the starting point of a simulation one needs the coordinates of the system and the initial velocities applied to it. The initial position of the atoms of the system is usually given by experimental data provided by, for example X-ray crystallographic structures. The initial velocities are randomly assigned to

each atom based on Maxwell-Boltzmann distribution, at the chosen temperature. The acceleration, which indicates how each particle moves from one position to another, is given by:

$$a = \frac{-1}{m}\frac{\partial V}{\partial r} \qquad \textbf{Equation 2.10}$$

where a is the acceleration, that can be calculated using the mass m of an atom in the position r, and the potential energy function V, that corresponds to the energy from the force field equation, mentioned above. The new positions are obtained using the acceleration as:

$$x = x_0 + tv_0 + at^2 \qquad \textbf{Equation 2.11}$$

being x the new position, t the used time step, v_0 the initial velocities and x_0 the initial position. The result of the successive application of these equations gives the trajectory of an atom for a given timescale. However, the potential energy and, consequently the acceleration, depends on all atoms in the system. As we deal with large systems as enzymes, this equation cannot be analytically solved, it can only be solved numerically [102]. Amber uses the *leap-frog* algorithm to solve these equations, in which the position and the accelerations of all atoms, at each time, is updated using the velocities of the previous step. Then, the new positions and accelerations will be used to calculate the next step [103].

2.2.2.3 Parameters to Run a Molecular Dynamics Simulation

Integration time step

The main goal of the MD simulations that we carry out is to explore the successive configurations of a system over time. To guarantee that the conformational space is effectively explored, the simulation needs to be sufficiently long. The choice of an adequate integration time step is also critical. If a large time step is used, the motion of molecules becomes unstable due to the significant errors in the integration of the equations of motion. Therefore, it can lead to errors on the molecular structure and velocities. In the other hand, if a very small time step is used, the calculation will not be efficient due to the very long calculation time. Our question here is what is the optimal length for a time step that could be able to give correct chemical information with computational efficiency? To prevent relevant physical errors, the time step in which the forces are calculated should be at least ten times smaller than the time of the fastest molecular vibration in the system. For simulating nuclear motions, as in MD simulations,

the fastest process is the motion of hydrogens, as they are the lightest particles. Hydrogen vibrations occur with a frequency of ~10^{-14} s^{-1}, and therefore a time step in the femtoseconds order (10^{-15} s^{-1}) is necessary to describe these motions [102]. This means that, for example, a simulation of only 1 nanosecond (10^{-9} s^{-1}) requires ~10^{6} time steps. As the stretching vibration of hydrogens have relatively little influence in the properties that we will measure with MD simulations, it is advantageous to fix all bonds that have these atoms. This approximation will allow for the use of longer time steps and, consequently, longer simulation times without increase the computational cost. The SHAKE algorithm is usually used to do this [104]. In biomolecules, the next faster vibrations correspond to C-C bond stretches, that occur with a frequency of ~ $2 \cdot 10^{-14}$ s^{-}, allowing a time step of 2 femtoseconds.

Solvent and Periodic boundary conditions

The properties of biomolecules are affected by the surrounding solvent through vdW and electrostatic interactions. Consequently, a MD simulation of enzymes must include a properly modelled solvent. In general, a large extend of explicit water molecules are added to solvate the solute (enzyme). The most typical solvent used with biomolecules are TIP3P or TIP4P, which have been parameterized to reproduce properties of water [105].

To be realistic, a model of a solution requires at least several hundred water molecules, which will increase the total system and, consequently, the computational cost. Furthermore, in a system with an enzyme surrounded by solvent, some of the water molecules will be on the solvent-vacuum interface, not interacting with the rest of the neighboring solvent molecules. To prevent the outer solvent molecules from distorting and minimizing surface effects, the periodic boundary conditions (PBC) are employed in MD simulations. As we refer above PBC is also a good strategy to prevent an infinite contribution from the explicit counting interactions.

In this method, the system (enzyme and water molecules) are placed in a unit cell, often having a cubic or octahedral geometry, that is infinitely repeated in every direction. If a molecule leaves the initial box, its image will enter, at the same time, from the neighboring box. This means that the number of molecules in the cell is constant during the MD simulation.

With infinite molecules (created by PBC) there are infinite contributions to the potential energy. The energies associated with bond stretching, angle bend and dihedral torsions are the same in all cells, being enough to calculate them for one cell. However, there are non-bonded interactions between molecules that are present in one cell and molecules present in the next repeated cells. Taking this in account, a *cut-off* distance is usually

employed to evaluating non-bonded interactions. A considerable saving in computational cost can be achieved if the vdW potential is truncated at some distance, usually ~10Å. For vdW interactions, this approximation does not affect, in large scale, the accuracy of the results. However, some artefacts are observed when it is applied to the electrostatic interactions. Contrary to the vdW energy, which falls of a r^{-6}, the Coulomb interactions fall with r^{-1}. A better method is to use two *cut-off* distances between which the E_{vdW} and E_{ee} decrease smoothly by a switching function. The more recent approaches use a third alternative, the Particle Mesh Ewald method [106,107].

Ensemble conditions

The macroscopic state that a MD simulation reproduces can be characterized by state functions, such as: number of particles (N), volume (V), temperature (T), pressure (P), total energy (E), and other properties. Only α+2 state functions have arbitrary values, all the other being defined univocally by the values chosen for the α+2 state functions (being α the number of components of system). The choice of the state functions whose value is defined depends on convenience and determine the "ensemble" of conformations generated by the corresponding MD simulation. There are many different ensemble conditions that could be used in a MD simulation. Three of the most common ensembles are: microcanonical (NVE), canonical (NVT) and isobaric-isothermal ensemble (NPT). In the NVE ensemble the number of particles, the volume of the box and the total energy of the system are kept constant during the simulation. In turn, in the NVT ensemble the number of particles, the volume of the system and the temperature of the system are maintained. On the other hand, NPT ensemble is characterized by a constant number of particles, and a constant pressure and temperature of the system during the simulation. In the initial steps of MD simulations, the canonical (NVT) ensemble is often used to maintain a constant volume of the system. Since the system was solvated with water molecules, it is important to use this ensemble to avoid excessive expansion due to the repulsive forces that still exist on the modelled system and that were not completely solved during the geometry optimization. After that, a NPT ensemble is usually used to continue the MD simulation.

Temperature and pressure control

The use of a thermostat is important during a MD simulation to preserve the temperature in a constant value through the simulation time. To control the temperature, the system may be coupled to a *"heat bath"*, which gradually adds or removes energy to/from the

system with an appropriate time constant (thermostat). The Langevin thermostat [108] is commonly applied in MD simulations. In Langevin dynamics the temperature is maintained by modified Newton's equations of motion, at each time step. Langevin equation can be used for MD equations by assuming that the simulated atoms are embedded in a large space with much smaller fictional particles. In many solvent-solute systems, the behavior of the solute (proteins or DNA, for example) is the desired one, and the behavior of the solute is not analyzed. In these cases, the solvent influences the solute by random collisions and, by imposing a frictional drag force on the motions of the solute in the solvent. These two effects are incorporated in Langevin equations

The pressure can similarly be held constant, or approximately constant, by coupling to a *"pressure bath",* using methods analogous to those used for maintaining the temperate [109].

2.2.2.4 Advantages and Pitfalls – summary

Molecular mechanics calculations are simple to perform and there are many software packages that allow this kind of calculations. In the works presented in this thesis, the AMBER 12 software package was used [110].

The big advantage of MD simulations is to allow the modelling of enormous molecules, such as enzymes. It is a primary tool of computational biochemistry.

The main disadvantage of MD simulations is that there are properties that one cannot measure with this method. Since chemical bonds are explicitly parameterized in the force field, it is not possible to describe chemical reactions that involve bond breaking and forming. Reactive force fields do exist, that parameterize the bonds with Morse potentials or similar, and that permit redefinition of the bonding on-the-fly, but they are very far from being accurate. Another important thing to mention is the MD sensitivity to parameterization. Due to this, the best manner to choosing a force field to use, with a specific system, is to check for similar systems/studies in the literature and also to compare the obtained results with experimental ones.

2.3 Quantum chemistry[2]

If we are interested in describing the chemical reaction that occurs in the active site of an enzyme, we need to describe the electron distribution in detail. Electrons cannot be

[2] The information summarized here was taken from General Books of Computational Chemistry as *Cramer et al*.; *Jensen et al.,* between others.

described in a classical way, being quantum mechanics essential to describe the electronic phenomena that occur at a subatomic scale. The general equation of quantum mechanics is the time-dependent Schrödinger equation:

$$\hat{H}\Psi(x,t) = i\hbar\frac{\partial}{\partial t}\Psi(x,t) \qquad \textbf{Equation 2.12}$$

In this expression i represents the imaginary unit, ħ corresponds to $\frac{h}{2\pi}$, t is the time, Ψ corresponds to the wave function and $\hat{H}$ is the Hamiltonian of the system. The wavefunction can be used to describe the position of the particles of the system (x) along the time (t), but, distinct to classical mechanics, that gives information about the exact positioning of the particles of the system, the wave function gives a probability amplitude. The square of Ψ describes the probability density of finding a particle of the system in a certain space and time.

The Hamiltonian ($\hat{H}$) is a mathematical operator that is applied to the wavefunction to allow the calculation of the total energy of the system. For a general N-particle system the Hamiltonian contains the sum of the kinetic energy ($\hat{T}$) of all particles and the potential ($\hat{V}$) energy for all nuclei (n) and electrons (e), as represented by Equation 2.13, which can be represented as in Equation 2.14:

$$\hat{H} = \hat{T} + \hat{V} = \hat{T}_n + \hat{T}_e + \hat{V}_{nn} + \hat{V}_{ne} + \hat{V}_{ee} \qquad \textbf{Equation 2.13}$$

$$\hat{H} = -\sum_K \frac{1}{2M_K}\nabla_K^2 - \sum_i \frac{1}{2}\nabla_i^2 + \sum_{K>L}\frac{Z_K Z_L}{|R_i - R_j|} - \sum_{i,K}\frac{Z_K}{|r_i - R_K|} + \sum_{i>j}\frac{1}{|r_i - r_j|} \qquad \textbf{Equation 2.14}$$

$$\nabla_i^2 = \left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}\right) \qquad \textbf{Equation 2.15}$$

where ∇_i^2 is the Laplacian operator (represented in Equation 2.15) acting on electron or nuclei, r and R are the distances between particles (electrons or nuclei).

For a N-particle system this equation is too complicated to be solved exactly. Due to this, in currently software used for the study of chemical reactivity this Hamiltonian is seldom used, being simplified by a series of approximations. Two approximations that are usually applied are:

1) the use of the time-independent Schrödinger equation;

2) the Born-Oppenheimer approximation.

In the time-independent Schrödinger equation it is assumed that the total energy of an isolated system is constant and not dependent on time (as Equation 2.12). The time independent Schrödinger equation can be represented as:

$$\hat{H}\Psi(x) = E\Psi(x) \qquad \textbf{Equation 2.16}$$

Being x the coordinates of all atoms in the system.

The other fundamental approximation is the Born-Oppenheimer approximation [111], that treats the electronic motions separately from nuclei motions, since nuclei move much slower that the electrons. In a practical point of view, the electronic relaxation with respect to nuclear motion is almost instantaneous. As such, the kinetic energy of nuclei can be ignored, and the nucleus/nucleus repulsion can be considered constant for a certain geometry, being the electronic energies computed for fixed nuclear positions. By considering stationary nuclei, and assuming that nuclear-nuclear repulsion can be added a posteriori and treated as a parameter, we can write:

$$\hat{H} = \sum_{i=1} \frac{-1}{2}\nabla_i^2 + \sum_{i=1}^{n}\sum_{K=1}^{N} \frac{-Z_K}{|r_i - R_K|} + \sum_{i=1}^{n}\sum_{i>j}^{n} \frac{1}{|r_i - r_j|} \qquad \textbf{Equation 2.17}$$

In which the first term is the sum over the kinetic energy of the electrons of the system, the second term corresponds to the interaction between electrons and nuclei and the third term is the sum of the interactions between pairs of electrons.

Born-Oppenheimer approximation leads to the concept of potential energy surface (PES). The PES corresponds to a surface that is defined by the electronic energy over all possible nuclear coordinates. The third term of the previous equation, that corresponds to the electrostatic interaction between electrons cannot be calculated analytically for any system with more than one electron.

To overcome this problem several methods to solve numerically the Schrödinger equation have been developed. The crudest ones are the semiempirical methods. The Hartree Fock method (HF) and a large family of post-HF methods provide numerical solutions with the accuracy that the user demands, at the cost of increasing the computational power. However, the numerical solution of the Schrödinger equation is not the pathway used in this work. Another branch of QM theory (Density Functional Theory) also allows to calculate the geometry and energy of molecules resorting to the electron density instead of the wavefunction. These methods are faster and more efficient, and we will describe them below.

2.3.1 Density Functional Theory

As seen above, the Schrödinger equation, even with the previous approximations is impossible to solve analytically. The density functional theory (DFT) is the more recent of the electronic structure methodologies and it has become an efficient alternative to Hartree Fock and post-Hartree-Fock methods. This is understandable due to its less computational cost when compared to other methods with similar accuracy. DFT attempts to address both the inaccuracy of HF theory and the high computational demands of the post-HF methods.

The idea behind DFT theory is that the energy of a system can be determined taking in account the electron density $\rho(\vec{r})$ at each point in space, instead of the complicated many-electron wave function. DFT only depends on three spatial variables ($\vec{r} = (x, y, z)$), independently of the number of electrons of the system. In contrast, the exact electronic wavefunction depends on three spatial and one spin variables for each electron of the system, which means a total of 4N variables.

DFT was proposed in 1964 by Hohenberg and Kohn and applied to finding the ground state electronic energy of a molecule. They showed that the ground state density of a system can defines all molecular properties of the system. It was also shown that this theory obeys the variational principle which shows that any calculated energy is always higher than the ground state energy. In DFT the electronic energy of the ground state of a molecule can be represented by different terms, with the density (ρ) as variable:

$$E_{tot}[\rho] = T[\rho] + E_{ee}[\rho] + E_{ne}[\rho] \qquad \textbf{Equation 2.18}$$

In this expression $T[\rho]$ represents the kinetic energy, $E_{ee}[\rho]$ corresponds to the electron-electron repulsion, and $E_{ne}[\rho]$ is the nuclear-electron attraction. The nuclear-nuclear repulsion is a constant within the Born-Oppenheimer approximation. The first two terms represent the density functional and are not dependent on the nuclear position:

$$F[\rho] = T[\rho] + E_{ee}[\rho] \qquad \textbf{Equation 2.19}$$

Taking this in account we can rewrite the previous equations as:

$$E_{tot}[\rho] = F[\rho] + E_{ne}[\rho] \qquad \textbf{Equation 2.20}$$

The $E_{ne}[\rho]$ corresponds to an external potential exercised by the nucleus in the system. The problem with the Hohenberg and Kohn formulation is that the exact form of the density functional $F[\rho]$ is unknow.

Kohn and Sham formulates lately a method that can give a practical solution to solve this problem. They introduce an orbital-based scheme similar to the HF methods. In this theory, the electron density is expressed as a linear combination of basis functions, called Kohn-Sham orbitals. This approach divides the kinetic energy of the system in two different parts. One is the kinetic energy of non-interacting system $T_S[\rho]$, with the same density as the real interacting system, and the rest corresponds to a residual kinetic energy, $T_C[\rho]$:

$$T[\rho] = T_S[\rho] + T_C[\rho] \qquad \textbf{Equation 2.21}$$

The functional of electron-electron repulsion, $E_{ee}[\rho]$ can be divided in a classical Coulombic interaction ($J[\rho]$) between the electrons and a residual non-classic term with the exchange, correlation and self-interaction ($E_{ncl}[\rho]$), as represented by Equation 2.22:

$$E_{ee}[\rho] = J[\rho] + E_{ncl}[\rho] \qquad \textbf{Equation 2.22}$$

Taking this in account the DFT total energy can be written as:

$$E_{tot}[\rho] = T_S[\rho] + T_C[\rho] + J[\rho] + E_{ncl}[\rho] + E_{ne}[\rho] \qquad \textbf{Equation 2.23}$$

Then the difference in the kinetic energy between the real system and the system with non-interacting electrons, $T_C[\rho]$, and the non-classic part of the electron-electron interaction, $E_{ncl}[\rho]$, can be combined to form an exchange-correlation functional ($E_{XC}[\rho]$):

$$E_{tot}[\rho] = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{XC}[\rho] \qquad \textbf{Equation 2.24}$$

The first three terms of Equation 2.24 can be calculated in an explicit way, however, the exchange-correlation functional is unknown and must me approximated to predict the total energy of the system.

To calculate DFT total energy it is necessary to use a self-consistent approach. This process allows to solve the above equations with an initial set of Kohn-Sham orbitals. Minimizing these equations results in equations that are similar to the HF method. This set of orbitals is used to calculate and improve the density. The process is iteratively

repeated, involving a variational process, until exchange-correlation energy term has satisfied some certain convergence criteria. If we know the exact value of $E_{XC}[\rho]$, the total energy of a system with many-electrons can be obtained. However, as this term is always an approximation, the DFT method results are dependent on how accurate this term is calculated. The biggest challenge of the DFT methods is centered on which functional to introduce to describe the $E_{XC}[\rho]$ term as exact as possible.

2.3.1.1 Exchange-correlation density functionals

In DFT, it is common to divide the exchange-correlation energy in two separated terms, as represented in Equation 2.25, although this approximation is not true since both the exchange and the correlation terms are contaminated with exchange-correlation.

$$E_{XC}[\rho] = E_X[\rho] + E_C[\rho] \qquad \textbf{Equation 2.25}$$

Many DFT approximations exist to calculate molecular properties at various levels of accuracy. The simplest approximation is based only in the electron density and it is called Local Density approximation (LDA). In this approximation the energy depends only on the density at the point where the functional is evaluated. It can be given by the density of electron in a homogeneous free electron gas. Taking this in account, an electron feels the electron density produced by the remaining electrons as if the density was the same in each part of the system. While this approximation works well for solid state systems, it does not result so good for systems in which the electronic density changes very much in space, as biological molecules. The first improvement to this approximation came with the creation of functionals that belongs to the so-called generalized gradient approximation (GGA). GGA functionals, in addition to incorporate the electron density, they take in account its gradient. Therefore, it becomes possible to describe inhomogeneous molecular densities in a better way. Examples of these functionals are BP86 [112] and PBE [113], which can be implemented efficiently and yield good results for structural parameters, however they are usually less accurate to describe other properties. More complex methods combine GGA functionals with part of HF calculations (usually the exchange integrals). These methods are called hybrid functionals. Nowadays, the predominant hybrid functional used to study biomolecular systems is the B3LYP functional [112,114,115], which has been employed in the study of enzyme catalytic mechanisms of the present thesis. A brief explanation of this functional will be given in the next topic. More recent theoretical methods include “meta-GGA” functionals, which extend the GGA corrections to more efficient methods and to “double hybrid” functionals. This functionals contains not

only the fraction of exchange from HF methods, but also a fraction of orbital-dependent nonlocal correlations energy estimated at the level of second-order many body perturbative theory (PT2).

2.3.1.2 B3LYP and other functionals

As mentioned above, B3LYP [112,114,115] is the most widely used functional to study biomolecules. Usually some parameters of the functionals are fitted to reproduce some set of observable properties. B3LYP is one of the Becke Three-Parameter Hybrid Functionals and its general form can be written as:

$$E_{XC}^{B3LYP} = (1-A)E_X^{Slater} + AE_X^{HF} + BE_X^{Becke} + CE_C^{LYP} + (1-C)E_C^{VWN} \quad \textbf{Equation 2.26}$$

In this expression, E_X^{Slater} is the Slater exchange, also referred to a Local Spin Density exchange, E_X^{HF} is the exact Hartree-Fock exchange, E_X^{Becke} represents the Becke's gradient of density correction of the exchange functional, E_C^{LYP} defines the LYP non -local correlation functional and E_C^{VWN} represents the VWN local correlation functional. A, B and C are constants determined by Becke via fitting a molecule set and have the values of $A = 0.20, B = 0.72$ and $C = 0.81$.

Many benchmarks have been done to test the performance of DTF functionals, especially for B3LYP accuracy. It is well documented that this functional has good accuracy on geometries. However, the studies on B3LYP accuracy on energy showed that this functional systematically underestimate the barriers, particularly in reactions that evolve transferences of heavy atoms. Unfortunately, there are no extensive benchmarks for enzymatic reactions, however Siegbahn and co-workers conclude that in general B3LYP gives an error around 3 kcal/mol in relative energies for enzyme reactions. In the case of systems involving transition metals, the error appears to be someway larger, but rarely above 5 kcal/mol [116].

There are also other types of functionals, as mentioned above. Inside de group of GGA functionals, which use derivatives (gradients) of the charge density, we have different types of functionals with different exchange and correlation terms. For example, BP96 (exchange: Becke and correlation: Perdew) or PW91 (exchange: pw91x and correlation: pw91c) [117], between many others.

Analyzing hybrid functionals group, in which the B3LYP functional is inserted, we can find many other functionals. The main difference between these functionals is the Hartree-

Fock exchange term. While B3LYP has 20% of HF exchange ($A = 0.20$), for example B1LYP [118] or mPW1PW [119] has 25% of HF exchange. During this thesis we also used Meta-Hybrid functionals (functionals that include the density, its gradient and the kinetic energy density, as well as HF exchange) to improve the energies of our calculations. Examples of this type of functionals are M06 and M06-2X, that contains 27% and 54% of HF exchange, respectively [120,121].

2.3.1.3 Limitations of DFT

DFT is an efficient and popular tool in computational (bio)chemistry. However, in addition to some of its advantages, pointed above (low computational costs, inclusion the correlation effects), it is important to remember that this method is not exact, and some limitations need to be pointed.

One of these limitations is related with self-interaction errors. In HF method the Coulomb repulsion between one electron and itself is cancelled by an exchange term. In DFT Coulombic terms are described exactly, but the exchange is described by an approximate functional. Since these terms do not cancel in an exactly way, there is a self-interaction error that remains. This error tends to decrease the calculated energetic barriers. Another error is related with the description of the wave function as a single determinant, which originates near-degeneracy errors, also called as errors in the non-dynamical correlation energy. Contrarily to the error associated with self-interaction, this error tends to increase the barriers and, because of that there is a substantial cancellation of the previous effect. Despite these limitations, some functionals, as B3LYP, are constructed to balance these errors, as well as possible. Another limitation of DFT functionals is the lack of description of van der Walls interactions (induced dipoles that result from electron correlated motions). It is also important to say that from system to system, or even from property to property, the performance of each functional does vary significantly, which is very undesirable [122].

2.3.1.4 Empirical dispersion corrections for DFT Calculations

The instantaneous charge fluctuations derived from the correlated motions of electrons gives rise to instantaneous dipoles that induced other dipoles and are responsible for dispersion energy (London dispersion). It has become very clear, especially for chemistry of large systems such as enzymes, that these interactions are very important to describe the energy of the system with accuracy.

These interactions are not correctly described by DFT methods beyond the very short-range, as pointed above, which compromises their applicability to systems where dispersion is important. DFT-D3 is an atom-pair wise dispersion correction proposed by Grimme and collaborators, that can be added to the Kohn and Sham DFT energies, as represented by Equation 2.27.

$$E_{DFT-D3} = E_{tot}^{DFT} + E_{disp} \qquad \textbf{Equation 2.27}$$

where E_{disp} corresponds to the dispersion energy, given by the sum of two and three-body energies:

$$E_{disp} = E^{(2)} + E^{(3)} \qquad \textbf{Equation 2.28}$$

This term is dominated by the two-body energy term, that is given by:

$$E_{disp} = \frac{-1}{2} \sum_{A \neq B} \sum_{n=6,8} s_n \frac{C_n^{AB}}{r_{AB}^n} f_{d,n}(r_{AB}) \qquad \textbf{Equation 2.29}$$

The first sum has in account all atom pair in the system. C_n^{AB} denotes the average *n*th order dispersion coefficient for atom pair AB, and r_{AB}^n is their internuclear distance. The s_n term is a scaling factor dependent on the functional used in the calculation. The value of the dispersion energy correction is frequently calculated separately from the DFT calculation. Its influence on the electron densities should be small, which allows a separate calculation and a subsequent addition to the DFT energy [123,124].

2.3.1.5 Basis set

In a quantum mechanical calculation, a basis set is a linear combination of mathematical functions that are used to represent atomic orbitals. The larger the basis set, the better description of the electron density obtained. However, the description of a complete basis set is not possible due to an infinite number of functions that are needed, but an adequate set of basis functions is enough to get good results. These mathematical functions that can be used to represent atomic orbitals can be divided into two principal forms:

- Slater type orbitals (STO);
- Gaussian type orbitals (GTO).

The STO are exponential functions. The exponential dependence on the distance between the nucleus and electron represents the exact orbital for hydrogen atom and it

ensures a rapid convergence with increasing numbers of functions. Nevertheless, the calculations of three or four-center are not possible analytically. These types of orbitals can be used for atomic and diatomic systems with a required high accuracy.

GTO are the most used type of basis set. The r^2 dependence makes GTO inferior to the STO. The GTO have problems representing the proper behavior near the nucleus. Furthermore, GTO falls off too rapidly far from the nucleus compared with STO orbitals. Taking this in account, more GTOs are necessary to achieve certain accuracy compared with STO's. Linear combination of n-GTOs (n = 3-6) can be used to reach a similar behavior to a STO. In terms of computational cost, GTO's are much more efficient because the product of GTOs in three/four centers can be calculated analytically and are usually used as basis functions in electronic structure calculations.

The next step is to choose the number of functions to be used. The smallest number of functions is called a minimum basis set. In this case, only enough functions are used to contain all electrons of the atom. For example, for hydrogen, a single *s*-function is adequate. For atoms of the first row of the periodic table it means two *s*-functions (1*s* and 2*s*) and a set of 2*p* functions ($2p_x$,$2p_y$,$2p_z$). The same logic is applied for the remained type of atoms.

The GTO can be improved combining a set of basis functions, known as the primitive GTO's (PGTOs) into a smaller set functions by forming linear combinations with fixed coefficients. The resulting functions are called contracted GTOs (CGTOs) These new functions improve the description of the atomic orbitals near the nucleus. There are many different contracted basis sets available in the literature. Pople style basis set are the ones used in this thesis and they will be briefly described below.

Pople Basis Set

A type of basis set created by Pople and co-workers are the *k-nlm*G basis set, which is a split valence type [97,125]. The *k* indicates how many PGTOs are used to describe the core orbitals of the system. The *nlm,* before the G, indicates the number of functions that the valence orbitals are split into, and how many PGTOs are used for their representation. If only two values are represented (*nl*), it indicates a split orbital and the basis set is called double-zeta. If three values are chosen (*nlm),* it represents a tripled split orbital (triple zeta orbitals). During this thesis we use one of the most used double-zeta basis set, that is 6-31G, to perform our calculations. In this basis set, the contracted core orbitals contain 6 primitives, and the valence electrons are represented by 2 basis, one with 3 contracted Gaussian functions (the inner part of the valence orbitals), and the other with only 1 (the

outer part of the valence orbitals). To improve these basis sets there are two special functions that are usually added to the previous ones: diffuse and polarization functions.
Diffuse functions are usually *s* and *p* functions, denoted by a + or ++ signal. The first + indicates one set of diffuse *s* and *p* functions on heavy atoms, and a second + represents a diffuse *s* function applied also to hydrogen atoms. Diffuse functions are important to improve the description of electronic densities that are spatially diffused and far from the nucleus.
In the case of anions, highly excited electronic states and large atoms, that tend to have more diffuse electronic density, some basis set do not have a good flexibility to allow a weakly bound electron to localize far from the electron density. In these cases, significant errors in energies and other molecular properties can be obtain, and therefore the use of diffusion functions are advisable.
Polarization functions are indicating after the G in the nomenclature, with a specific designation for heavy atoms and for hydrogens. These functions are added to provide some flexibility to the plain orbitals. For example, if the analyzed orbital is a *s* orbital, by adding polarization functions the size of the orbital can be changed to adapt better to the molecular environment and to be polarized by it. For example, the 6-31G(d) basis set is a split valence basis, with a single *d*-type polarization function on heavy atoms. This is the basis set that we use in the optimization calculations during this thesis. The 6-311++G(2d,2p) is also used during this thesis to improve the energies of the enzymatic barriers. It corresponds to a triple split valence basis set, with 1 set of diffuse *s* and *p* functions on heavy atoms, a diffuse *s* applied to hydrogen atoms, and polarization functions (two *d* functions on heavy atoms and two *p* functions on hydrogen atoms).
All these options show that, in addition to the choice of the *ab initio* method to use in the calculations of biological molecules, a correct choice of the basis set is also an essential step.

2.4 Hybrid methods

2.4.1 Introduction

Until here, we saw that only small systems can be treated with quantum mechanics. However, the curiosity and the necessity to study the chemistry of complex molecular systems (such as enzymes) and the catalytic principles underline their reactions, leads to the formulation of the hybrid quantum mechanics and molecular mechanics methods. Combining quantum mechanics and molecular mechanics (QM/MM) methods are an essential tool in modern computational enzymology. The use of QM methods to treat the

active site residues of enzymes, allows the accurate description of the reactions catalyzed by enzymes. More than that, these hybrid schemes allow the inclusion of the surroundings/environment effects, treating the residues outside the active site with MM methods. With QM/MM methods the atomistic description of the enzyme mechanisms and the identification and characterization of reactive species, transition states (TSs) and reaction intermediates, is now possible.

The work of many groups has been influential in the development of the hybrid methods and their applications. The study of Warshel and Levitt of lysozyme was one of the firsts in this area [126].

Taking in account the main topic of this dissertation - the catalytic power of enzymes – this type of methods is the first choice to try to explore this topic, as they allow the atomistic study of enzyme catalysis. As we list in the first chapter of this dissertation, there are many controversial proposals looking for explain the origin of enzyme catalysis, including the role of enzyme dynamics in catalysis, the search of “near attack conformations”, entropic effects and so on. QM/MM methods can help to provide evidences to support these proposals. Many reviews have documented the development of the QM/MM approach and its application to biomolecular systems [88,127-131].

In QM/MM methods the total energy of the whole system (denoted as *real system*), defined as $E_{QM/MM}$, is represented as a sum of the energy of the system treated by QM method (denoted as *model*), represented as E_{QM}, the energy of the environment treated by the MM methods (E_{MM}) and the interaction between the QM *model* system and the MM environment system (E_{QM-MM}):

$$E_{QM/MM,real} = E_{QM,model} + E_{MM,environment} + E_{QM-MM} \qquad \textbf{Equation 2.30}$$

This scheme is known as an “additive scheme”. The energy from the interaction between QM and MM system is difficult to calculate. It generally includes: 1) bonded interactions for covalent bond(s) that are located between the QM and MM boundary (stretching, bending, torsional contributions) and 2) non-bonded interactions (van der Walls and electrostatic interactions). In this case, all interactions between layers are evaluated in a classic way.

In turn, the QM/MM energy can also be described as a “subtractive” scheme. The ONIOM method (our own N-layer Integrated Molecular Orbital and Molecular mechanics) is a hybrid method based in a subtractive scheme [88,129,132,133]. In the ONIOM method the system can be divided in two, or more, layers, which can be treated with different or equal methods (QM/MM or QM/QM). In the next subchapter we will discuss with more detail the ONIOM subtractive method, which was used in the works presented in this thesis: 1) to

describe the catalytic mechanism of human renin; 2) to describe the catalytic mechanism of PatGmac; 3) to study the influence of frozen residues on the QM/MM methods; 4) to study the influence of a buried water on α-amylase catalytic mechanism and 5) to study the influence of different structures on the catalytic mechanism of HIV-1 protease.

2.4.2 Subtractive ONIOM scheme

The ONIOM method, developed by Morokuma and coworkers [132], is a subtractive or extrapolative method. In this method the total energy of the entire system is given by the QM energy of the *model* system, that includes the reactive atoms ($E_{QM,model}$), plus the MM energy of the total *real* system ($E_{MM,real}$), minus the MM energy of the *model* system ($E_{MM,model}$). The expression that represents the ONIOM energy is defined as:

$$E_{ONIOM} = E_{QM,model} + E_{MM,real} - E_{MM,model} \qquad \textbf{Equation 2.31}$$

The "double-counted" of the *model* system atoms is removed by the subtractive term. This subtraction, of the MM energy of the *model* system from the $E_{MM,real}$ of the entire system ($E_{MM,real} - E_{MM,model}$), gives the energetic contribution of the environment on the *model* system, calculated by an inexpensive method. In the original ONIOM scheme, all QM/MM interactions are treated using the MM force field, which is referred to as a mechanical embedding scheme (ME or ONIOM-ME). Later, an electronic embedding scheme (EE or ONIOM-EE) was also developed, allowing the polarization of the QM wave functions by the atomic charges of the MM part. In the next section we will discuss these two schemes. Compared to the additive scheme, the ONIOM method does not need the additional coupling Hamiltonian (third term of Equation 2.30). Looking at the ONIOM formulation (Equation 2.31), one notice that the MM calculations of the *real* and *model* systems are performed, and, because of that, a good set of parameters are required to obtain a reliable energy. Actually, only the MM atomic charges and the parameters for QM/MM boundary are important, due to cancellation of the classical MM terms in the ONIOM scheme.

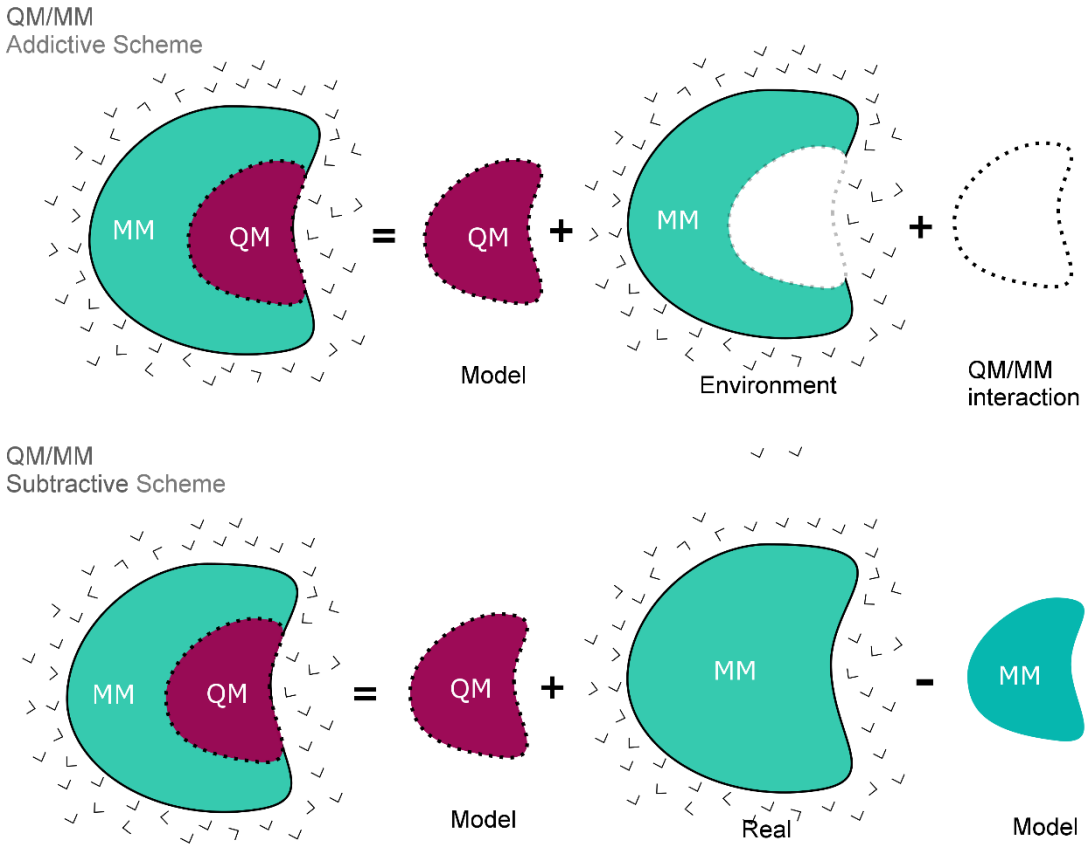


Figure 2.2 Comparison of different layers division used in the QM/MM additive scheme and in the QM/MM subtractive scheme (ONIOM scheme). QM and MM layers are denoted by violet and blue, respectively.

2.4.3 Electrostatic Interactions between layers

As mentioned earlier, there are different forms to handle the interaction between the QM layer and the MM layer. They differ between them by the degree of mutual polarization between the QM and MM regions. Here we will discuss the mechanical and electrostatic embedding schemes.

Mechanical Embedding

In mechanical embedding scheme (ME) the electrostatic interaction between MM and QM layer is included in the second term of the ONIOM energy equation (Equation 2.31), which means that it is evaluated at the MM level of theory. In other words, those interactions are described as classical Coulomb interactions between point charges. This method is straightforward, computationally efficient and depends in the atomic charges assigned to the *real* and *model* system. The simplest way to calculate this energy is to use the default

charges defined within the force field, no matter how much the geometry of the *model* system changes during the optimization steps. However, the electronic state and, consequently, the charge distribution, can change during the geometry optimization and, the energy is dependent on these changes. When large charge transfer occurs during a reaction, the use of the initial set of charges on the *model* system may not provide an accurate result. This scheme will often not be accurate enough in the case of very polar environments and it is important to mention that in biomolecules the environment around the active site have usually polar residues or water molecules. In these cases, the atomic charges of the *model* system may be recalculated during the study of the reaction, using, for examples an electrostatic potential (ESP) or restrained electrostatic potential (RESP) charges to fitting the electrostatic potential around the *model* system [134]*.* In **Chapter 4** (The study of the catalytic mechanism of Renin) a detailed methodology using this protocol is described).

ONIOM-ME scheme is computationally very efficient, nevertheless, it has some also limitations:

1) the charges in the MM region do not interact with the QM density which is not influenced by the electrostatic environment. Thus, the *model* system charge density is not polarized by the MM charges[3];

2) the charge distribution in the QM region needs to be updated during the study of the reaction, as the charge distribution of the QM model changes during the reaction;

Electrostatic Embedding

The major limitation (the lack of interaction between the charges of the QM and the MM regions) of the ME scheme can be eliminated by taking in account the presence of MM charge model on the QM calculation. This procedure will account for the polarization of the *model* system under the perturbation exerted by the point charges of the environment system. The energy of the ONIOM electrostatic method is approximated as:

$$E_{ONIOM-EE} = E^{V}_{QM,model} + E_{MM,real} - E^{V}_{MM,model} \qquad \textbf{Equation 2.32}$$

This equation is very similar to the E_{ONIOM} presented above; however, in this case, the first and the third term are calculated in a different way to include the effect of the

[3] Although this effect should be small, as showed by us in HIV-1 protease (please see **Chapter 4** – pag. 158)

polarization of the *model* system in the QM calculation. The first term is calculated by incorporated the MM point charges in the QM Hamiltonian and it can be represented in as in Equation 2.33:

$$H^{V}_{QM,model} = H_{QM,model} - \sum_{i}\sum_{A_{env}} \frac{s_{A_{env}} q_{A_{env}}}{r_{iA_{env}}} + \sum_{A_{QM}}\sum_{A_{env}} \frac{Z_{A_{QM}} s_{A_{env}} q_{A_{env}}}{r_{A_{QM}A_{env}}} \qquad \textbf{Equation 2.33}$$

In this expression q represents a point charge in the environment system, r is the distance between a *model* system's electron, or nucleus, and a point charge in the environment, subscript QM refers the model system and *env* to the environment system. Here, two additional terms are applied to the QM Hamiltonian. The extra terms of the Equation 2.33 account for the polarization of the environment on the model system. When the QM layer is treated with a DFT method (the most used method to treat the model system) those terms are added to the model one-electron operator of the Kohn-Sham equations.

It is also important to say that a scale factor is used in the second and third term. This scale factor is usually 1 (which means that no scaling is performed) for most of the MM point charges, and it could be 0 for the atoms near the boundary. If all the point charges from the environment system are included in the QM Hamiltonian, the model system seems to suffer over-polarization. Therefore, the point charges, typically for atoms that are within three bonds away from the last atom in the model system are excluded from this calculation by using a scale factor of 0.

Beside the electrostatic interactions there are other important type of nonbonded interactions: the van der Waals (vdW) interactions. In the EE scheme, the vdW interactions are calculated in the same way as in ONIOM-ME scheme.

Using EE scheme, the electronic structure of the model system can adapt to the environment, and the electronic QM density should be much closer to the reality, than when it is calculated in gas-phase (ONIOM-ME).

We mentioned above that in ONIOM-EE scheme, the electrostatic interactions are calculated within the QM Hamiltonian and the point charges of the nearby atoms are zeroed out. It is important to mentioned that the contribution of these atoms is calculated in the second term of the **Equation 2.32**, thus their contributions are not totally ignored.

Nowadays, most of the QM/MM work uses EE scheme. In this dissertation, we apply the ME scheme in our first work (**Chapter 4**) and then we change to EE scheme and applied it in the next works.

2.4.4 QM/MM Boundary treatment

When applying QM/MM ONIOM methods to enzyme catalysis, a few covalent bonds need to be cut between the layers, which leads to the generation of unsaturated bonds. Because of that, the boundary between the *model* system and the environment needs to be treated with care. Several different methods have been used to lead with this problematic of cap the dangling bonds. The approach adopted in ONIOM methods involve link atoms. Link atoms schedule introduces an additional atomic center which is typically a hydrogen (H-link) atom, that is placed between the atom of the *real* system and the atom of the *model* system, covalently bonded to the last one. This new atom is included in the *model* system, but it is not part of the *real* system. It is a common procedure to fix the line of the link atom in such that it remains in the same position of the bond that was cut, with a well-defined distance from the atom of the QM layer that is obtained by multiplying the original bond by a constant factor.

2.4.5 QM/MM geometry optimization

In studies with small molecules, the location of the relevant stationary points (minima and transition states) are commonly performed exploring potential energy surfaces (PES). In QM/MM studies of large molecules, as biomolecule, that have thousands of atoms, it is also possible to optimize relevant stationary points. In these cases, one needs techniques to handle efficiently thousands of atoms (many degrees of freedom). There are different algorithms capable of manipulating coordinates and optimizing energies taking advantage of the partitioning of the system into different layers. Among different strategies we will only mention the microiterative optimization strategy, that is the one applied in the ONIOM method. In this strategy the geometry of the core region (containing the QM region) is relaxed alternately with the environment. This alternation allows the efficient optimization of minima and transition states in large molecules even when electrostatic embedding is used.

In this scheme, the geometry of the environment system is fully optimized at a given *model* system geometry. After achieving the zero forces for the environment system, the atoms present in the *model* system move one step forward (macroiteration). This procedure is iterated until the system reaches a stationary point on the potential energy surface. In Gaussian 09 [135] (the software used in the calculations performed in this thesis), at every geometry optimization step, the program evaluates if the calculated is converged or not. Forces and displacements have a particular importance in this evaluation. The optimization is regarded as converged if these values are smaller than threshold values.

There is a vast configuration space that is accessible in biomolecules during the QM/MM calculations, there are many closely minima and transition states for the same reaction. Due to that, the exploration of the stationary points along a single reaction path is usually not recommended. Sometimes, it is suggested to determine at least several representative transition states and minima, starting with different initial structures (different snapshots from classical MD simulations can be used as starting structures), to assess the diversity of the environment conformational space and its influence.
In this thesis, this procedure was applied to the rate-limiting reaction of HIV-1 protease and α-amylase and fluctuations on the reaction barriers were observed.

2.4.6 Pros and Cons of ONIOM method

The QM/MM applications are by now considered as the best powerful computational technique to treat reactivity in large systems. They can be applied to model an electronic event, in the active site of biomolecules, hundreds of atoms treated with quantum mechanics, and considering the influence of the environment. However, one of the biggest pitfall of ONIOM method is that its reliability is strongly dependent on the choice of the low and high methods, and on the preparation of the *model* and *real* systems. Errors from an improper preparation of the system are very common but very difficult to identify and correct.
It is also important to have in mind that the exploration of the PES of a single enzymatic reaction is dependent on the initial protein configuration, as was reported by many previous works, and was also observed during this thesis. Therefore, the proper preparation of the system is essential, and multiple starting structures are recommended.
It is also highly recommended to decompose the total energy of the system (E_{ONIOM}) into the QM and MM contributions, to understand the energetic contribution of each layer. The large contribution to the energy should come from the high layer (where the electronic reaction occurs), however the energy that comes from the low layer also helps to understand how the environment contributes to the reaction and provided the steric strain that limits the conformations of the *model* system to a realistic range.
In the next section, we describe a possible protocol to perform ONIOM QM/MM calculations. Nevertheless, it is only a suggestion between other possibilities.

3.5 How to model an enzyme catalytic mechanism?

3.5.1 A possible protocol

Preparation of the initial enzyme-substrate system

ONIOM method is frequently used to study enzymatic reactions as pointed above. When an input file for ONIOM calculation is prepared, the protocol to study an enzymatic reaction can be routinely performed, just by following some steps in which some theories and procedures described above are applied. One of the most important and time-consuming step corresponds to the preparation of the system/model before any calculation.

To study a reaction mechanism of an enzyme, one needs to start with the heavy-atom coordinates of the protein, which are often available in the literature, from X-ray crystallographic or NMR studies. One of the best sources for search and download the coordinates of an enzyme is the Protein Databank library available at www.rcsb.org, and known as Protein Data Bank (PDB) [136]. The coordinates of these crystal structures are extremely useful as initial points for modelling enzyme reactions, however the user should analyze them in detail, and check if there are missing atoms/residues or even mutated residues. Hydrogen atoms are not easy determined by X-ray diffraction and thus they are not included in the PDB files.

Because of these points, after the choice of the PDB file, it is crucial to add the hydrogen atoms to the structure, editing it and remove some crystal buffer additives. Usually, the hydrogen atoms can be directly determined considering the hybridization of the atoms, however the protonation states of the titratable residues must be done with care. There are different tools that help to predict the local pKa values of each residue inside a protein, such as PROPKA [80,81,137,138] and H++ [139]. This automated protonation of the enzyme residues does not dispense the visual inspection by the user. A wrong protonation of the system can compromise the correct study of the reaction mechanism.

Due to the rapid turnover of enzymes it is very difficult to crystalize them with the natural substrate. Thus, the modelling of the substrate in the active site is also an important step. Sometimes, the enzyme is crystalized with an analogue of the substrate and, in these cases, small modifications and superimpositions with other similar structures are enough to model the natural substrate in an adequate position to promote the reaction. In other cases, there is a need to construct and dock the substrates in the active sites. Sometimes, the natural substrate is present in the active site, however some enzyme residues are mutated.

When the substrates are organic molecules, their charges, vdW parameters and parameters for intramolecular interaction need to be determined, in contrast to amino acid

residues and common cofactors/biomolecules, for which these parameters have already been determined and are present in the force-field. The final PDB structure, before starting the calculations, must have the substrate in the correct pose in the final PDB file.

The resultant protein-substrate structure is then solvated and minimized (since the obtained geometry may not correspond to a minimum in the force field, and to prevent possible clashes in the system) before any QM/MM or MD simulation.

Sometimes, the minimized structure is used directly in QM/MM calculations, however, even though, as mentioned above, enzymes have many degrees of freedom and therefore, using a single structure to study catalysis cannot be enough to understand a catalytic mechanism and its origin. On the other hand, using potential energies, as the one calculated by ONIOM equation for measuring the relative stability of different states does not necessarily provide sufficient information about the reactivity. Another solution is to calculate the free energies, considering a large number of possible conformations of the complex enzyme-substrate. In this case, to generate conformational sample for free energy calculations, MD combining with QM/MM is the method of choice. However, combining these two methodologies using an accurate QM level could lead to large or even prohibitively computational demands. This type of calculations becomes possible if some approximations are made to reduce the computational cost, as the use of semi-empirical calculation instead of DFT, to treat the QM layer of the system.

In the works presented in this thesis, and in other similar works, another strategy was adopted. A MD simulation was performed and used to produce different snapshots of enzyme-substrate complex with different geometries. The study of the same reaction mechanism starting with different snapshots are useful to understand how different residues, in distinct positions, could influence the catalytic mechanism of an enzyme. This protocol gives a more reliable picture of the reaction pathway when compared to direct ONIOM calculation using only the minimized crystal structure.

After the preparation of the initial model (docking the substrate and performed molecular dynamics, if necessary), it is important to make sure that the connectivity of the system is defined correctly for all atoms. If we performed pure QM calculation, it is not necessary to define the connectivity, because this information is not used when solving the Kohn and Sham equations. However, in ONIOM calculations it is required to indicate how all atoms in the system are connected. Then, the next step is the choice of the *model* and the *real* systems.

The simulation system

Now, before the QM/MM calculation, the system is ready to be divided in different layers to start the mechanistic study. At this point, there are two important questions to having in mind:

1) What residues one should include in the QM layer (*model* system)?
2) How to cut them?

The QM layer must include, not only all atoms that are directly involved in the chemical reaction, but also a significant part of the catalytic residues and substrate. More than that, an adequate choice of the high layer must ensure that the intramolecular interactions established around the active site are also included. This means that hydrogen bonds, π-interactions and ionic bonds formed between the reactive and the nearest atoms must be included in the *model* system. This procedure seems to be simple, however the compromise between a representative QM layer and the computational cost makes this task more difficult. For enzymatic catalysis, when DFT is used to treat the *model* system, it is now possible to use a QM layer with over one hundred atoms. DFT layers with more atoms may be very demanding computationally, however small layers may not take in account some important effects from the surroundings.

There are some other considerations to take in account in the definition of the QM layers:

1) Atoms linked by double or triple bonds should be placed within the same ONIOM layer (or QM or MM);
2) Region boundaries should not fall within aromatic rings, or delocalized double bonds;
3) The molecular mechanics parameters should be the same within the real and model systems for all stationary points.

In the MM layer the cleanest strategy is to include the whole protein plus a hydration shell. The inclusion of a hydration shell is highly recommended, to shield the polar residues present at the protein surface. In the crystal model, only the water molecules occupying a well-defined position are present, because the mobile ones are not easily detectable. One option to solvate the system is use a sufficiently sized sphere of water molecules around the protein. This protocol should guarantee that the surface of the protein is adequately hydrated. When this cap of water molecules is not used it is advisable to fix the surface residues of the protein to prevent self-interactions and biased results.

The next step is to choose the size of the region which will be free to move during the optimization steps, while the remaining system would be fixed or restrained. The convergence of the geometry optimization tends to become problematic if many atoms are free. The effect of restraining some atoms during the QM/MM optimizations was evaluated during this thesis and it is described in **Chapter 6** (The Influence of frozen residues in the exploration of PES of enzyme reaction mechanisms). A reasonable and very used selection for the free region includes all atoms within 15Å far from the QM layer, however we observed that free regions with above 6Å far from the QM atoms seems to be enough, to obtain reliable results.

Choice the QM and MM Methods

As mentioned above, DFT is currently the preferable level of theory to perform QM calculations in QM/MM models. As also mentioned, B3LYP functional is a safe choice for geometry optimizations, and to explore the potential energy of reaction coordinates [71,116,140] . The energy is more sensitive to the functional choice and, consequently, single point energy calculations, using the optimized structures, are usually performed using different functionals. There are different benchmarking studies performed for several reaction properties that could help in the choice of the density functional. In the MM layer, the choice of the force field depends more on the alternatives available in the software. In the works performed in this thesis we used the force field ff99SB [141,142].

Reaction Coordinate exploration

When the enzyme-substrate complex is prepared, the next step is optimized it with QM/MM methods (minimize its energy). This step may take some time to achieve the convergence criteria. Sometimes, when the convergence fails one can solve this using different initial structures collected from different frames from the MM simulation, as mentioned before. Other solutions are: i) use different levels of theories (starting for example with mechanical embedding instead of electrostatic embedding scheme); ii) use different functionals or basis set.

Having an optimized enzyme-substrate complex, it becomes possible to start the exploration of possible reaction coordinates to describe the reactions that takes place. Using this structure, linear scans are performed to explore the potential energy surfaces (PES) along the reaction coordinate. Typically, bond lengths are used as reaction coordinate, but we can also define angles or dihedrals. The initial set of possible reaction

coordinates usually derived from experimental and computational hypothesis and chemical intuition [143].

A linear scan of a good reaction coordinate will have a maximum in energy and a minimum relatively to all other coordinates. The maximum can be used as a guess to find the transition state for the explored transformation. The optimization of transition states requires an initial calculation of the force constants that will drive the optimization of the structure to a maximum. This step might be computationally demanding. When the structure of the transition state is optimized, an IRC (intrinsic reaction coordinate) can be conducted. IRC corresponds to the minimum energy reaction pathway in mass-weighted cartesian coordinates between the transition state of a reaction and its reactant and product. The obtained minima for reactant and product / intermediate may not coincide with the initial ones, obtained with the linear scan along the reaction coordinate.

The final reaction profile and thermochemistry quantities

To establish the final reaction profile, it is common to perform a thermodynamic characterization of all steps, to obtain free-energy profiles wish are more comparable to experimental data. The thermochemistry data is not obtained directly from the calculations that we described until now and the calculated energies do not correspond to free energy. Only the electronic energy is calculated during the QM/MM optimizations, and the kinetic energy due to the nuclear motions is often discarded. However, to compare computational results with experimental ones, it is advisable to include the energy associated with these motions in the final energy. These motions persist at 0K, and the internal energy of the system at this temperature (E_0), is equivalent to the electronic energy (E_{elec}), which includes the repulsion energy of the nuclei and the zero-point energy (ZPE):

$$E_0 = E_{elec} + ZPE$$ **Equation 2.34**

A Thermal correction is added to the computed energy in order to obtain energies at the selected temperature. These thermal corrections are determined taking in account the contribution from translation, rotation and vibration of the nuclei. Thus, the final energy, at a selected temperature (E_{temp}), is given by the sum of the internal thermal energy correction and the internal energy (E_0).

$$E_{temp} = E_0 + (E_{translational} + E_{rotational} + E_{vibrational})_{temp}$$ **Equation 2.35**

Frequency calculations provides the vibrational frequency for each vibrational mode of a molecule. When the stationary point corresponds to a transition state, one of those

frequencies is imaginary, while if the structure is a minimum all frequencies have positive values. These calculations are highly demanding in computational resources that are available and on the size of the system. A frequency job must be calculated using the same theoretical model and basis set as produced the optimized geometry

Sometimes, it is useful to compute the Gibbs free energy and to that it is necessary to taking in account the entropy of the system. To calculate the entropy, it is assumed that only the fundamental electronic state is occupied and, because of that the electronic entropy of the non-degenerated system is zero. Consequently, the entropy is computed only considered the translational, rotational and vibrational contributions of the nuclei, using similar approximations that are used to complete the thermal corrections, as described before [144,145]. In systems with frozen coordinates it is not advisable to include the values of zero-point energy and the vibrational thermal entropy and enthalpy in the final energies. These values are not mathematical correct because the minima obtained dos not corresponds to a real stationary point, and, therefore, these values cannot be applied with in a rigorous way [94].

Generally, as mentioned above, energy single point calculations with different functionals, and more complete basis set, are also performed in the optimized geometries. Empirical dispersion corrections, as Grimme D3, may also be introduced in the final energies to describe better classical long-range interactions.

2.6 Other QM/MM and QM methods to study enzyme catalysis

In this thesis we described and worked only with QM/MM methods based on single conformation or multiple conformation exploration of PES with geometry optimization with ONIOM. However, there are a large diversity of methods that can be applied to study enzymatic catalysis. Here, we will only provide a brief presentation of the most common methods: QM/MM MD (Car-Parrinelo MD or semi-empirical MD simulations), EVB (Empirical Valence Bond) method and Cluster Model.

2.6.1 QM/MM MD

The QM/MM energy can be in principle be used within any MD simulations. The main objective of such simulation is the sampling of configuration space. Approximate treatments have been developing to reduce the computational cost of this calculations. Both molecular dynamics and Monte Carlo methods are used to perform sampling. In

most cases semi-empirical methods are routinely applied in QM/MM MD, to treat the QM layer [57,146-148]. Nevertheless, there are examples for QM/MM MD that uses DFT or Hatree-Fock (HF) in the calculations of the high layer.

Another option regarding time-dependent methods is Car-Parrinelo Molecular Dynamics (CPMD) [149]. This method simultaneous calculate the nuclei and electron movements from DFT calculations.

2.6.2 EVB

Warshel and co-workers developed a different scheme in which the sampling is performed using an empirical valence bond (EVB) potential fitted to ab initio data [150-152]. This method is a QM method (that can be implemented within a QM/MM scheme) used a valence bond approach to describe enzymatic reactions. The system wavefunction is characterized by a linear combination of the most important ionic and covalent resonance forms. The electronic Hamiltonian is built using parameters extracted from empirical values and/or from *ab initio* surfaces. This method allows the determination of the free-energy profiles of chemical and enzymatic reactions with low computational cost, even in system with a large number of atoms. This is possible because, EVB approach is built as a sum of parametric functions very similar to MM force field without any explicit treatment of electrons, which makes this method many times fasters than QM/MM methods. This is the main advantage of this method, which have demonstrated very good quantitative results when compared with experiments, provided that the incorporated empirical terms are cautiously chosen. However, the main disadvantage of this method is that with the choose of the valence bond forms, one implicitly chooses how the direction of the chemical reaction should be.

2.6.3 Cluster Model

Another alternative to study enzymatic catalysis is the cluster model approach. This method does not use two different methodologies as the previous ones. Here, the system contains a limited number of atoms (mostly two-three hundred), and all the atoms are treated with QM methods [153].

The most important issue in cluster model techniques is the choice of the model. Generally, all the residues that are involved in the reaction and the substrate (or part of the substrate) are included. More than that, all the residues that can influence catalysis through the formation of short and long-range interactions within the reactive part, should

be included. Nowadays, it is possible to apply this methodology to about 300-400 atoms in the limit. This selection is typically treated with DFT, or other ab initio method. A dielectric constant is generally used to include long-range non-specific effects.
The main disadvantage of this method is its applicability in systems for which long-range interactions by specific regions far from the active site play a fundamental role in catalysis, or in system in where significant conformational changes may have influence in catalysis.

2.7 Conclusions

Computational chemistry has become an essential tool to study biomolecules and, more specific to describe the catalytic mechanism of enzymes. The large part of the success of this recent chemistry field comes from the increasing of the computational power during the last years, as well as, to the constant development of computational /theoretical methods described above. Some of these methods, described in this chapter in the user perspective and in a simple way, are mathematically complex, and can offer a detailed, atomic resolution insight of the behavior of biomolecules, which cannot be obtained by experimental results.
One of the ultimate goals of computational chemistry applied to enzymology is to be able to describe the entire system (the enzyme-substrate complex substrate and solvent) with only QM methods. Although it is not possible do obtain this yet, the field of computational enzymology is still growing, and continuous improvements are pursued.
The methodologies described here were applied in the next chapters, with exception for the **Chapter 3** that was a more statistically work with no computational chemistry calculations. In the other ones we applied or single structure QM/MM methodology (**Chapter 4** and **Chapter 5**) or QM/MM methodology using different initial structures (**Chapter 6, Chapter 7 and Chapter 8**).

CHAPTER 3. Activation free energy, substrate binding and enzyme efficiency fall in a very narrow range of values for all enzymes

Sérgio Filipe Sousa[a], Ana Rita Calixto[a], Maria João Ramos[a], Carmay Lim[b] and Pedro Alexandrino Fernandes[a]

[a] UCIBIO, REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal, [b] Institute of Biomedical Sciences, Academia Sinica, Taipei 11529, Taiwan, Department of Chemistry, National Tsing Hua University, Hsinchu 300, Taiwan

With this chapter, we start the presentation of the works developed during the last four years. One of the first objectives, that was defined for this thesis, was to collect available kinetic and mechanistic information for a large number of enzymes and correlate this information to identify possible patterns responsible for enzyme efficiency. In the following manuscript, we presented a systematic analysis of different enzyme parameter among several enzymes, focus in differences and similarities between each group of enzymes presented in **Chapter 1**. Although this paper is not yet published, we decided to start with it, to give an overview of all enzyme's classes and their catalytic power.

Regarding the contributions to the paper, Ana Rita Calixto analyzed all the information collected by Sérgio F. Sousa and wrote the first draft manuscript, which is prepared to submit.

3.1 Abstract

Enzymes are responsible for controlling many biological processes, catalyzing different reactions using substrates with diverse complexity, size and chemistry. Why enzymes are so efficient and what is the origin of their catalytic power are some of the fundamental questions of biochemistry. The present work consists in a systematic analysis of three parameters directly related with enzyme catalysis (activation free energy, substrate binding free energy and enzyme efficiency) and their correlations with structural and biological information, considering the presence/absence of different cofactors, the oligomerization state, their size (the number of residues by monomer), the cellular location, temperature and the substrate specificity/promiscuity. These parameters were estimated using the turnover number and the K_M value for each enzyme at the corresponding temperature. They are useful to understand differences/similarities between different types of enzymes and they could be important to elucidate the origin of the catalytic power of enzymes. The results show that, regardless of the large diversity of enzymes and their different chemical reactions, the values of activation free energy, substrate binding free energies and enzyme efficiencies fall in a very characteristic narrow range of values for all enzymes. The strategies to achieve such performance are different in each enzyme, being the presence of cofactors, their location, their size or oligomeric state, their specificity or even the environmental conditions essential to confine the values of the catalytic parameters to a narrow range.

KEYWORDS

Enzymes, catalytic power, binding free energy, activation free energy, enzyme efficiency

3.2 Introduction

Most chemical reactions that occur in life are facilitated by enzymes. Understanding the origin of their catalytic power remains one of the big tasks of enzymology, which has not only a fundamental, but also practical interest and impact. This topic has been one of the central questions of biochemistry. However, like other fundamental questions, the origin of the enzyme catalytic power continues to be a matter of debate, despite the large quantity of structural data that is presently available.

There are many reviews that discuss how enzymes work [2,13,20,27,154]. The Linus Pauling hypothesis [155] that affirms that enzymes lower the activation energy of biologic reactions by binding transition states better than substrates is now accepted, however, how enzymes do that is still a non-consensus question. Looking for an answer to this query is not an easy task. This is because there is a vast number of enzymes with many structural differences, with different sizes, that use a huge number of different substrates and catalyze many different reactions. During the last years, many different hypotheses have been suggested to explain enzyme rate enhancements. A list of these hypotheses include **i)** desolvation [31], **ii)** reduction of binding entropy [40,156], **iii)** orbital steering [157], **iv)** low barrier hydrogen bonds [158], **v)** dynamic effects [13,81,159], **vi)** tunneling [154], **vii)** electrostatic preorganization of the active site [20], **viii)** preorganization of the enzyme active site residues to hold the substrate in a reactive conformation (near attack conformations – NACs) [160], **ix)** change of the acid/base pK_a [161,162], and **x)** nonequilibrium specific vibrations [163], in between others.

In the present paper, we focus on a global analysis of enzymes' properties in all the six classes defined by The International Union of Biochemistry and Molecular Biology (IUBMB), trying to identify some features that could help to understand how enzymes work. How do these different enzymes, which catalyze different reactions, lower the energy of biological reactions? Have they similar strategies to decrease the energy associated to the reactions that they catalyze?

In this work, we present a detailed analysis of the behavior of three parameters directly linked to the catalytic power of enzymes: **i)** substrate binding free energy (ΔG_{bind}), **ii)** activation free energy ($\Delta G^{\ddagger}_{cat}$) and enzyme efficiency ($\Delta G^{\ddagger}$).

In one of our previous papers we have shown that hydrolases' values for ΔG_{bind}, $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}$ fall in a rather narrow range, despite the large diversity of substrates upon which they act on [164]. This contrasts markedly with the rates of biological reactions in absence of a catalyst that vary over at least 19 orders of magnitude, some with half-lives with more than one million years, as reported by Wolfenden [12].

Here, we analyze the catalytic data available for the different classes of enzymes trying to identify differences and similarities between them. Moreover, the analyzed parameters (ΔG_{bind}, $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}$) are correlated with other enzyme properties, such as their size and oligomerization state, the type of reactions that they catalyze, the presence and type of cofactors, their cellular/extracellular location and substrate specificity and environmental conditions, all for the benefit of trying to identify some trends and correlations that could help to understand the catalytic power of enzymes.

3.3 Methodology

3.3.1 Collecting data on enzymes

For collecting data on each enzyme class, BRENDA [165,166] (BRaunschweig ENzyme DAtabase) was used to extract information about different enzymes taking into account some specific criteria:

1) four well-defined EC numbers;

2) available experimental conditions (ideal temperature and pH);

3) available functional parameters, as turnover number and K_M.

Information on enzymes lacking these criteria or with unspecific experimental conditions or mutant species were excluded.

The BRENDA enzyme database contains information on enzymatic reactions with a large variety of possible substrates, a number that often includes molecules that are non-native substrates. Although enzymes are able to catalyze some reactions involving such substrates, sometimes in artificial conditions, these typically occur with lower reaction rates. For proper analysis of the enzymatic reaction trends, it is therefore critical to verify that the enzyme parameters used in the analysis correspond only to native substrates. Accordingly, each individual substrate in the initial data set was carefully verified taking into consideration the known enzyme-substrate specificity of the corresponding enzyme and the current knowledge reported in the enzyme annotations in the ExPASy resource portal [167,168].

This information was complemented and checked with information available in the most recent literature for each enzyme. A final manually curated database containing a total of 934 enzymes with representatives from the 6 classes defined by the IUBMB and with confirmed physiological substrates and associated enzyme parameters, was obtained.

3.3.2 Enzyme properties analyzed

In this work, we analyzed three parameters directly related with enzymatic catalytic power: binding free energy (ΔG_{bind}), activation free energy ($\Delta G^{\ddagger}_{cat}$) and enzyme efficiency ($\Delta G^{\ddagger}$). Since the turnover number is equal to the first order rate constant, k_{cat}, when the enzyme is saturated with substrate, regardless of the nature and identity of the rate–limiting step, it was used (collected from BRENDA) to approximate the k_{cat}. K_M values were collected directly from BRENDA. The enzyme's catalytic efficiency was estimated as k_{cat}/K_M. The K_M, k_{cat}, and k_{cat}/K_M, values and the corresponding temperatures were used to determine ΔG_{bind}, $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}$ [19] (**see Figure 3.1),** as previously done for hydrolases [164].

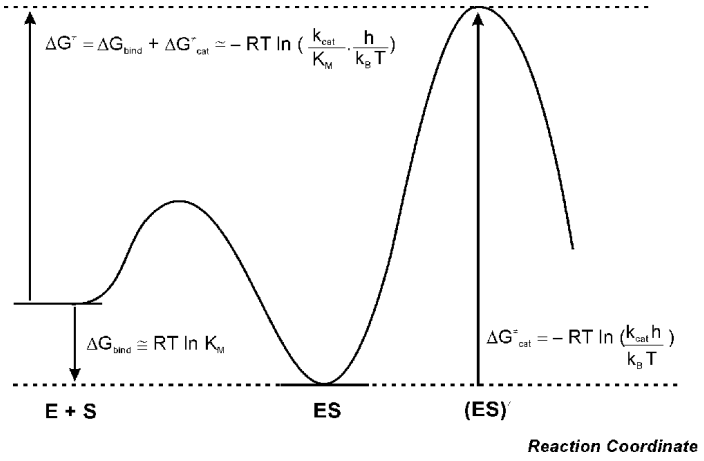


Figure 3.1 Schematic representation of the energetics of enzyme reactions, showing the relationship between K_M, k_{cat} and k_{cat}/K_M, values and ΔG_{bind}, $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}$. $\Delta G^{\ddagger}$ represents a measure of the catalytic efficiency of enzymes, being related with the second-order rate constant, relevant for enzymes and substrates at physiological concentration; ΔG_{bind} corresponds to the binding free energy, and $\Delta G^{\ddagger}_{cat}$ corresponds to the apparent activation free energy for the first-order reaction when the substrate is present at saturating concentrations.

The kinetic information for each enzyme in the database was manually complemented with structural and biological information retrieved not only from BRENDA [165,166] but also from the UniProtKB/Swiss-Prot databases [169]. Among the features, were the cellular location of the specific enzyme isoform, oligomerization state, the number of amino acid residues for each monomer, and the type of cofactor(s) that are present. Each BRENDA entry was additionally classified as acting on a single specific substrate or on more than one substrate (promiscuous enzymes). A process of manual verification of the information present in BRENDA and in the UniProtKB/Swiss-Prot databases, was also performed. This verification process checked the uniformity of the information included for each entry and species, including number of amino acid residues, number of chains in the wild-type active enzyme, native metal cofactor and substrate preference. This verification was

based on: **(i)** the corresponding references in the literature, **(ii)** available structures in the Protein Data Bank and their references, and **(iii)** the amino acid sequences of the enzyme in the UniProtKB database. This methodology was achieved to validate the data, correct possible inconsistencies and minimize errors in the data set. The same methodology was used in our previous work for hydrolases [164].

3.3.3 Statistical analysis

For each class with more than 20 entries (each entry corresponds to one enzyme, with a specific substrate, from a specific organism, within specific reaction conditions), the average value of each enzyme parameter as a function of a given enzyme class was computed. To compare different classes/groups and determine if the obtained average values are statistically different, a two–tailed *t*–test was performed. The two-tailed t-test gives the probability *p* that any two datasets are not statistically different, considering the size of each dataset, the distribution of the corresponding values (in this case the values of each enzyme parameter for each enzyme) and their average value (the average value of each parameter in a given enzyme class). The confidence level that the values derived from two datasets are statistically different is given by $(1 - p) \times 100$. The average values derived from two different datasets are considered different when the probability is computed to be less than 5%.

It is important to mention that this work was based on experimental data present in the BRENDA database, which can have much information for the same enzyme in different experimental conditions or expressed in different species. This large amount of information could possibly result in few incorrect or incoherent data, despite the fact that all entries were manually inspected and checked to minimized possible errors.

3.4 Results

3.4.1 Variations in the enzyme parameters (ΔG_{bind}, $\Delta G^{\ddagger}_{Cat}$ and $\Delta G^{\ddagger}$)

The results presented in **Table 3.1** show the average substrate-binding free energies (ΔG_{bind}), apparent activation free energies ($\Delta G^{\ddagger}_{cat}$) and the catalytic efficiency values ($\Delta G^{\ddagger}$), for all 934 enzyme entries, as well as for the different classes of enzymes, together with the corresponding standard deviation and the number of entries for each category. **Figure 3.2** illustrates the distribution of the same parameters for all entries and for each class of enzymes.

Table 3.1 Average values of ΔG_{bind}, $\Delta G^{\ddagger}_{cat}$, and $\Delta G^{\ddagger}$ for each class of enzymes and for all enzyme entries. Despite the diversity, the values of ΔG_{bind}, $\Delta G^{\ddagger}_{cat}$, and $\Delta G^{\ddagger}$ are similar between all enzyme classes.

Class	no. of enzymes	ΔG_{bind} (kcal/mol)	$\Delta G^{\ddagger}_{cat}$ (kcal/mol)	$\Delta G^{\ddagger}$ (kcal/mol)
EC.1. Oxidoreductases	214	–5.7 ± 1.8	15.7 ± 1.7	10.0 ± 2.0
EC.2. Transferases	164	–5.8 ± 1.9	17.2 ± 2.0	11.5 ± 2.2
EC.3. Hydrolases	339	–5.5 ± 1.9	16.6 ± 2.2	11.0 ± 2.4
EC.4. Lyases	82	–5.6 ± 1.7	16.5 ± 1.9	11.0 ± 2.3
EC.5. Isomerases	82	–5.0 ± 1.8	15.2 ± 2.4	10.2 ± 2.7
EC6. Ligases	53	–5.8 ± 1.5	17.4 ± 1.6	11.6 ± 2.0
All Enzymes	**934**	**–5.6 ± 1.8**	**16.4 ± 2.1**	**10.8 ± 2.4**

3.4.1.1 Binding free energy, ΔG_{bind}

There are many experimental and theoretical studies with evidences that the reduction of the free energy of activation by enzymes can be explained by the enzyme/substrate interactions. Could the energy related to the binding of the reactants (ΔG_{bind}) provide some answers or clues on the origin of enzymes' catalytic power? The data presented in the third column of **Table 3.1** and **Figure 3.2 (panel A)** could help to answer these questions. Despite the extremely diversified set of enzymes included in this study, and the diversity of the corresponding substrates and type of reactions catalyzed, the results show that the ΔG_{bind} values are remarkably similar for different enzymatic classes, falling in a very limited range of values. The average ΔG_{bind} value for the full set of enzymes (934 entries) was of -5.6 kcal/mol, with a standard deviation of only 1.8 kcal/mol. Comparing different classes, the average binding free energy values range between -5.8 kcal/mol (transferases and ligases) and -5.0 kcal/mol (isomerases).

Figure 3.2 shows that 85% of all enzymes have substrate-binding free energies in a narrow range between -3 to -7 kcal/mol, being the peak at -5 kcal/mol (21%). This result is in perfect agreement with our previous report concerning the ΔG_{bind} distribution in hydrolases [164], showing that this particular parameter seems to be independent of enzyme class.

To further make sure of this, we also analyzed the distribution of the ΔG_{bind} in each individual enzyme class. The results show that transferases (EC. 2) and ligases (EC. 6), have the lowest (i.e. most negative) average ΔG_{bind} values (**Table 3.1** and **Figure 3.2**, panel A, purple and grey lines, respectively). However, it is important to notice that the average ΔG_{bind} values for oxidoreductases (EC. 1, -5.7 kcal/mol), hydrolases (EC. 3, -5.5

kcal/mol) and lyases (EC. 4, -5.6 kcal/mol) were not found to be statistically different from these. Isomerases (EC. 5), which catalyze changes within one molecule, have, on the other hand, a less negative average ΔG_{bind} (-5.0 kcal/mol) statistically different from all other classes (with more than 95% of confidence level). While all other classes' peaks are between -6 or -5 kcal/mol, isomerases have a maximum distribution at -4 kcal/mol.

In isomerases, the initial reactant and the final product are very similar (isomers). Having this in account, if the ΔG_{bind} is more negative, making the substrate binding stronger, the release of the product would be more difficult. Therefore, these less negative values of ΔG_{bind} for isomerases are very interesting and logical, having in account the particularities of the reaction catalyzed by this class of enzymes. The results show that despite some differences among the ΔG_{bind} values, these parameter range between -3 and -7 kcal/mol, independently of each class enzymes belong to.

These global results contrast with other observations of binding free energies between larger and smaller molecules in nature. In particular, Zhang & Houk [27] have described a distribution of binding free energies for diverse types of host-guest complexes that varied between +3 and –15 kcal/mol. Smith & co-workers [170] have also described a wider distribution of binding free energy values when addressing a collection of protein-ligand complexes from the Binding MOAD database [171] . The authors have shown a significant distribution of values in the range –3 to –15 kcal/mol, with some ligands having ΔG_{bind} values as large as –20 kcal/mol.

Summarizing, our results show that ΔG_{bind} values are similar for different enzymatic classes and are limited to a small range of values, excluding isomerases, that present less negative values for this parameter.

3.4.1.2 Activation free energy, $\Delta G^{\ddagger}_{cat}$

The BRENDA database has information for k_{cat} values, that can be converted into activation free energies, $\Delta G^{\ddagger}_{cat}$. The catalytic power of enzymes can be related with a decrease of the activation free energy of the catalyzed reaction. Have all enzymes similar activation free energies or could this parameter be characteristic of each class of enzymes?

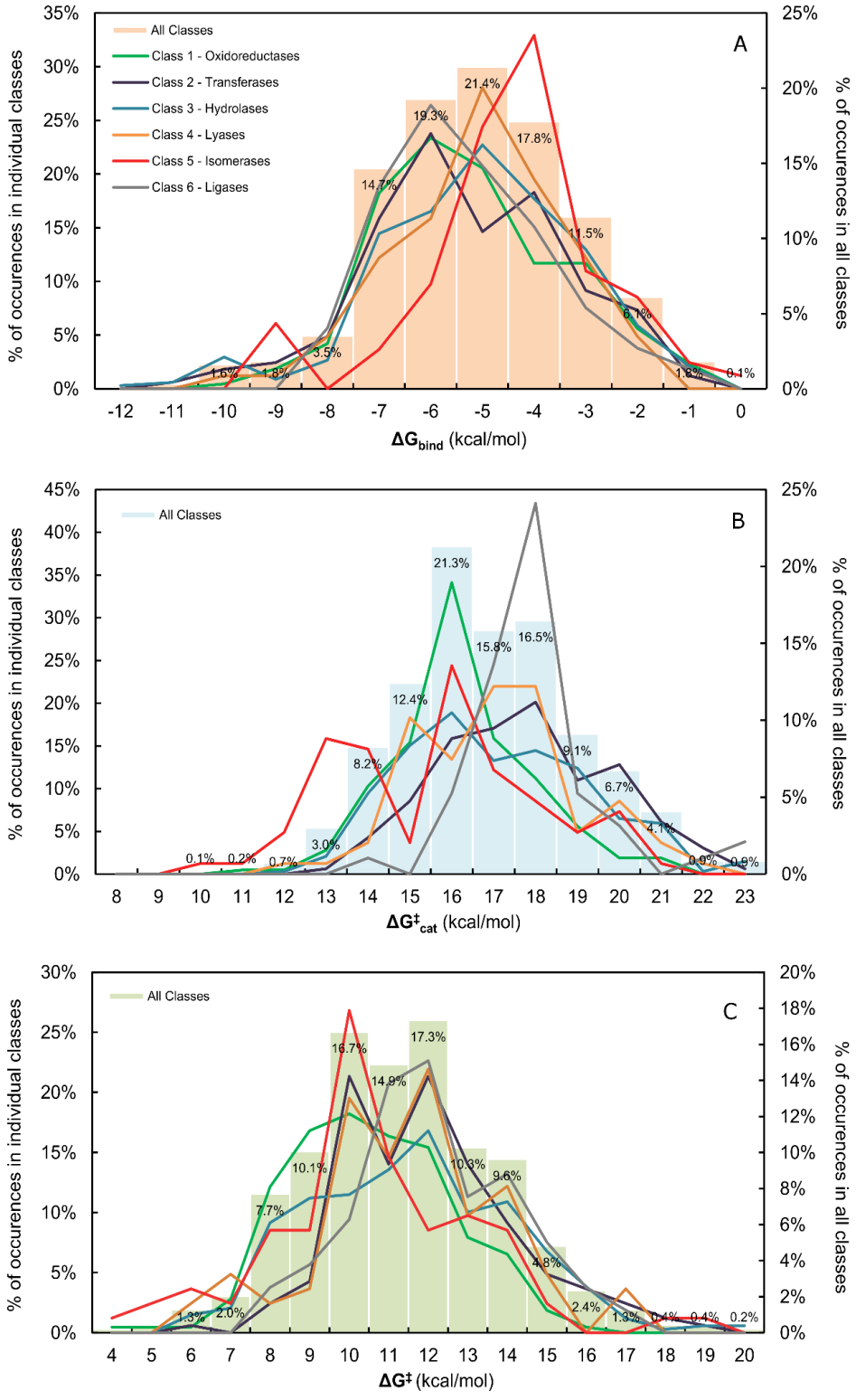


Figure 3.2 Distribution of ΔG_{bind}, $\Delta G^{\ddagger}_{cat}$, and $\Delta G^{\ddagger}$ (panels A, B and C, respectively) among enzymes.The % of occurrence frequencies for all enzymes are shown as histograms, while those for the individual classes are shown as curves in different colors. Despite the diversity the values of ΔG_{bind}, $\Delta G^{\ddagger}_{cat}$, and $\Delta G^{\ddagger}$ fall in a very narrow range for all enzymes.

The results presented in **Figure 3.2** show that apparent activation free energy ($\Delta G^{\ddagger}_{cat}$) values for 94% of enzyme entries fall in the range between 14 and 21 kcal/mol. 66% of all

values fall on a narrow range of 15-18 kcal/mol. The average $\Delta G^{\ddagger}_{cat}$ is 16.4 ± 2.1 kcal/mol (**Table 3.1**). The standard deviation of 2.1 kcal/mol is small, particularly taking into consideration the extremely large structural diversity of the enzymes considered, which span between all six classes of enzymes, involving extremely diverse sets of substrates and reactions. Once more, the results agree with previous results for $\Delta G^{\ddagger}_{cat}$ distribution in hydrolases [164].

Analyzing the results for each class (**Table 3.1**) we notice that transferases (EC.2) and ligases (EC. 6) have statistically-relevant higher average activation free energies (17.2 and 17.4 kcal/mol, respectively) than the other classes (more than 95% of confidence level), which is indicative of slower reaction kinetics. On the other hand, oxidoreductases (EC.1) and isomerases (EC.5) have statistically significant lower average $\Delta G^{\ddagger}_{cat}$ values (15.7 and 15.2 kcal/mol), corresponding to faster kinetics (more than 99% of confidence level compared to other classes). Hydrolases (E.C 3) and lyases (EC. 4) have intermediate average $\Delta G^{\ddagger}_{cat}$ values (16.6 and 16.5 kcal/mol). The distribution of $\Delta G^{\ddagger}_{cat}$ values for the different enzymatic classes (lines in **Figure 3.2**) agrees with the tendencies noticed in the analysis of the average $\Delta G^{\ddagger}_{cat}$ values by class (**Table 3.1**) - oxidoreductases, hydrolases, lyases and isomerases (EC. 1, EC. 3, EC. 4 and EC. 5) have a maximum distribution at 16 kcal/mol, while transferases and ligases (EC. 2 and EC. 6) show peak at 18 kcal/mol.

3.4.1.3 Enzyme efficiency, $\Delta G^{\ddagger}$

The catalytic efficiency estimated as a sum of ΔG_{bind} and $\Delta G^{\ddagger}_{cat}$, ranges from 8 to 14 kcal/mol for 87% of all enzymes (**Figure 3.2C**) with an average $\Delta G^{\ddagger}$ of 10.8 ± 2.4 kcal/mol (**Table 3.1**). The more efficient enzymes are oxidoreductases (EC.1) and isomerases (EC.5) with average $\Delta G^{\ddagger}$ values and maximum distributions around 10 kcal/mol, being significantly different from the other classes (confidence levels between 96% and ~100%), statically speaking.

Transferases (EC.2) and ligases (EC.6) have, in turn, similar enzymatic efficiency values, being less efficient than the other classes (11.5 and 11.6 kcal/mol). Transferases and ligases are enzymes that need two molecules to undergo the necessary catalysis reactions Transferases catalyze the transfer of a functional group from one molecule to another and ligases are responsible for joining two different substrates. This need for two substrates to work, could be one possible explanation for their smaller efficiency. Interestingly, these two classes have the higher $\Delta G^{\ddagger}_{cat}$, but the more negative ΔG_{bind}, as described above, meaning that they bind the substrates very well but subsequently are very slow in performing the chemical transformations.

3.4.2 Correlation between $\Delta G^{\ddagger}$ and $\Delta G^{\ddagger}_{cat}$ or ΔG_{bind}

$\Delta G^{\ddagger}$ was estimated as a sum of ΔG_{bind} and $\Delta G^{\ddagger}_{cat}$, as represented in **Figure 3.1**, but nevertheless it is important to understand which one of these two quantities has more influence in the catalytic efficiency. Taking that into account, we plot $\Delta G^{\ddagger}_{cat}$ and ΔG_{bind} in function of $\Delta G^{\ddagger}$ for all entries of enzymes and for different classes. We proceeded to correlate them by a linear regression Accordingly, both the slopes and r^2 were calculated for each plot and are shown in **Table 3.2** and **Figure 3.3**. The slopes of the adjusted lines show how the values of $\Delta G^{\ddagger}_{cat}$ and ΔG_{bind} vary with $\Delta G^{\ddagger}$. Large slopes indicate that $\Delta G^{\ddagger}$ is more sensitive to evaluated parameter. The value of r^2 indicates how the values fit with the line of the regression. Looking at all entries, the catalytic efficiency ($\Delta G^{\ddagger}$) seems more dependent on the activation free energy ($\Delta G^{\ddagger}_{cat}$) (**Figure 3.3B**) than on binding free energy (ΔG_{bind}) (**Figure 3.3A**), as shown by the slopes of the lines adjusted to the values (0.60 and 0.40, respectively). Looking at each individual class the results are similar (**Table 3.2**) with only the oxidoreductases having the inverse tendency. In this case the efficiency seems to be a little bit more sensitive to the substrate binding free energy (slope 0.55) than to the activation free energy (slope 0.45). The individual analysis of each class shows that there is no correlation between the values of $\Delta G^{\ddagger}_{cat}$ and ΔG_{bind}, as shown by the small r^2, obtained from the adjusted curve. Interestingly, the value of $\Delta G_{bind}/\Delta G^{\ddagger}_{cat}$ is very similar between all classes of enzymes. In 83% of all entries this value ranges from 0.6 and 0.8.

Nji9m s

Table 3.2 Correlation between $\Delta G^{\ddagger}$ and $\Delta G^{\ddagger}_{cat}$ or ΔG_{bind} for all enzymes and for each class of enzymes.

		ΔG_{bind} vs $\Delta G^{\ddagger}$		$\Delta G^{\ddagger}_{cat}$ vs $\Delta G^{\ddagger}$		$\Delta G^{\ddagger}_{cat}$ vs G_{bind}	
Classe	**no. of enzymes**	**Slope**	**r^2**	**Slope**	**r^2**	**Slope**	**r^2**
EC.1. Oxidoreductases	214	0.55	0.40	0.45	0.30	-0.32	0.09
EC.2. Transferases	163	0.47	0.28	0.53	0.34	-0.36	0.14
EC.3. Hydrolases	339	0.41	0.28	0.59	0.44	-0.24	0.06
EC.4. Lyases	82	0.40	0.31	0.60	0.50	-0.17	0.04
EC.5. Isomerases	82	0.32	0.22	0.68	0.55	-0.18	0.06
EC.6. Ligases	53	0.44	0.34	0.56	0.45	-0.20	0.05
All Enzymes	**934**	**0.40**	**0.27**	**0.59**	**0.44**	**-0.25**	**0.08**

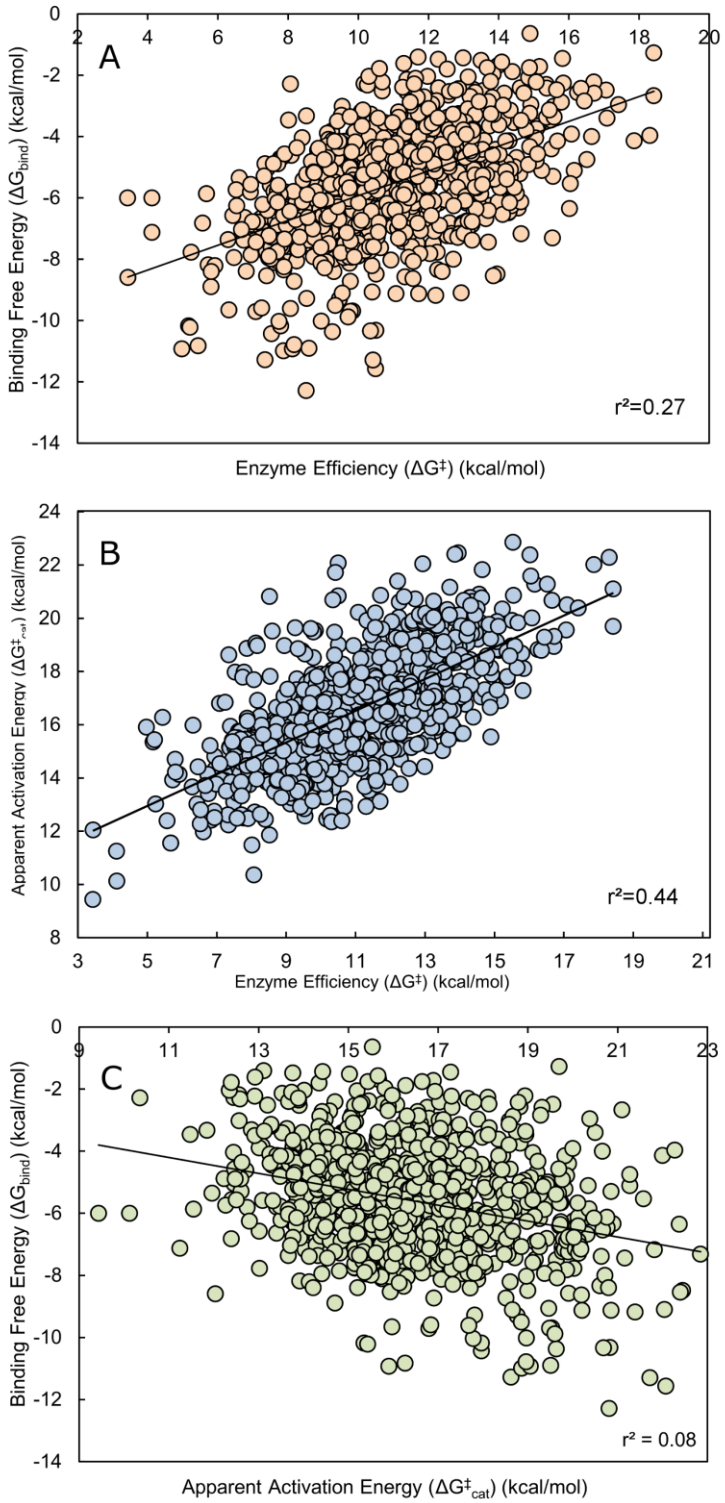


Figure 3.3 Correlation between $\Delta G^{\ddagger}$ and ΔG_{bind} or $\Delta G^{\ddagger}_{cat}$, and between ΔG_{bind} and $\Delta G^{\ddagger}_{cat}$ for all enzymes. Panel A represents the correlation between ΔG_{bind} and $\Delta G^{\ddagger}$, panel B represents the correlation between $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}$. In both cases the correlation is small, but the results indicate that $\Delta G^{\ddagger}$ is more sensitive to $\Delta G^{\ddagger}_{cat}$ than to ΔG_{bind}.

3.4.3 Dependence of enzyme parameters on cofactors

Many enzymes are dependent on small molecules, called cofactors, to perform their reactions. Cofactors act on the enzyme structural stabilization or even in the catalysis, participating directly in the catalyzed reaction [172-174]. Therefore, a relation between cofactors and enzyme parameters ($\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and consequently $\Delta G^{\ddagger}$) is expectable.
Among all analyzed entries, there are 620 that use cofactors (inorganic or organic), and 314 that do not contain any cofactor to catalyze their reactions (or without information about cofactors in the initial database), as represented in **Table 3.3.**
The average value of $\Delta G^{\ddagger}_{cat}$ and ΔG_{bind} is lower for enzymes that employ cofactors compared to those that do not use them (16.3 vs. 16.6 kcal/mol, 88.8% confidence level and -5.6 vs -5.5 kcal/mol, 80.9% confidence level). The efficiency of the two types of enzymes is different within statistical relevance, being the enzymes that employ cofactors more efficient than those that do not have cofactors (10.7 vs 11.1 kcal/mol, 98.5% confidence level) – see **Table 3.3**. However, despite being statistically relevant, the differences are very small.

Table 3.3. Average $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$ for enzymes that employ cofactors and for those which do not, and respective standard deviations. The results have been divided by class of enzymes and are represented in kcal/mol.

Classes	Cofactor	no. Enzymes	ΔG_{bind}	$\Delta G^{\ddagger}_{cat}$	$\Delta G^{\ddagger}$
EC.1. Oxidorreductases	With	214	-5.7 ± 1.8	15.7 ± 1.7	10.0 ± 2.0
	Without	0	-	-	-
EC.2. Transferases	With	93	-5.6 ± 1.8	17.1 ± 2.1	11.5 ± 2.0
	Without	71	-6.0 ± 2.0	17.4 ± 1.9	11.4 ± 2.4
EC.3. Hydrolases	With	180	-5.8 ± 2.0	16.3 ± 2.1	10.5 ± 2.3
	Without	159	-5.3 ± 1.7	16.9 ± 2.1	11.6 ± 2.5
EC.4. Lyases	With	44	-5.3 ± 1.6	17.1 ± 1.8	11.8 ± 2.0
	Without	38	-5.8 ± 1.7	15.9 ± 1.9	10.1 ± 2.3
EC.5. Isomerases	With	36	-5.0 ± 1.8	15.8 ± 2.4	10.8 ± 3.4
	Without	46	-5.0 ± 1.9	14.7 ± 2.4	9.6 ± 1.9
EC.6. Ligases	With	53	-5.8 ± 1.5	17.4 ± 1.6	11.6 ± 2.0
	Without	0	-	-	-
all enzymes	With	**620**	**-5.6 ± 1.8**	**16.3 ± 2.0**	**10.7 ± 2.5**
	Without	**314**	**-5.5 ± 1.8**	**16.6 ± 2.3**	**11.1 ± 2.3**

Looking at each class of enzymes though, the results are slightly different. First, there are two classes which use, in most cases, organic cofactors to catalyze the corresponding reactions: oxidoreductases (E.C. 1) and ligases (E.C. 6). Oxidoreductases (E.C 1) are

responsible for catalyzing the transfer of electrons from one molecule to another, using the organic cofactors to do that. In the case of ligases (E.C. 6), they catalyze the addition of two molecules by forming a new chemical bond. This kind of enzymes uses nucleoside triphosphates, diphosphates or monophosphates in their corresponding reactions.

In each of these two classes, there are some entries that, in addition to the organic cofactor, use also metals as cofactors. These cases will be included later on in this paper. As far as the other three classes are concerned, it is possible to compare the previous parameters between enzymes that use cofactors and enzymes that do not use them. Starting by transferases (E.C. 2), among the 164 entries, there are 93 that employ cofactors (mostly metals), whereas 71 do not employ them in their reactions (or at least there is no information on that).

Considering $\Delta G^{\ddagger}_{cat}$, its average value is smaller in transferases which employ cofactors (17.1 and 17.4 kcal/mol, 72% confidence level). However, the value of the binding free energies (ΔG_{bind}) is less negative (less favorable) when a cofactor is present (-5.6 and -6.0 kcal/mol, 85% confidence level). In this particular class, the efficiency of the enzymes is similar in both cases (11.5 and 11.4 $kcal.mol^{-1}$).

In the hydrolases (E.C 3), as showed in our previous work [164], the number of enzymes that employs cofactors is very similar to the one that does not (180 with cofactor, 159 without cofactor). In this group, enzymes that employ cofactors have smaller $\Delta G^{\ddagger}_{cat}$ than enzymes that do not (16.3 and 16.9 kcal/mol, 99.4% confidence level), which is a similar result to that observed in the E.C. 2 class of enzymes. Comparing the values for ΔG_{bind}, we find that the average value is more negative for enzymes that employ cofactors, meaning that substrate binding to these enzymes is more favorable (-5.8 and -5.3 kcal/mol, 98.3% confidence level). These results are contrary to that observed in the E.C. 2 class of enzymes. Additionally, the average values for $\Delta G^{\ddagger}$ show that hydrolases that use cofactors also have higher efficiency compared to those that do not employ cofactors (10.5 and 11.6 kcal/mol, ~100% confidence level).

In class 4 (Lyases) the tendency is different - the average activation energy $\Delta G^{\ddagger}_{cat}$ is higher in enzymes that use cofactors (17.1 and 15.9 kcal/mol, 99.6% confidence level) and the values of ΔG_{bind}, are less negative for lyases that employ cofactors (-5.3 and -5.8 kcal/mol, 74.9% confidence level). Contrary to the previous analyzed classes, lyases with cofactors seem to be less efficient than enzymes without them (11.8 and 10.1 kcal/mol, 99.9% confidence level).

The results observed in class 5 (Isomerases) show that enzymes with cofactors have on average lower activation free energies (14.7 and 15.8 kcal/mol, 95.6% confidence level), however the values of ΔG_{bind} are very similar between them (-5.0 and -5.0 kcal/mol).

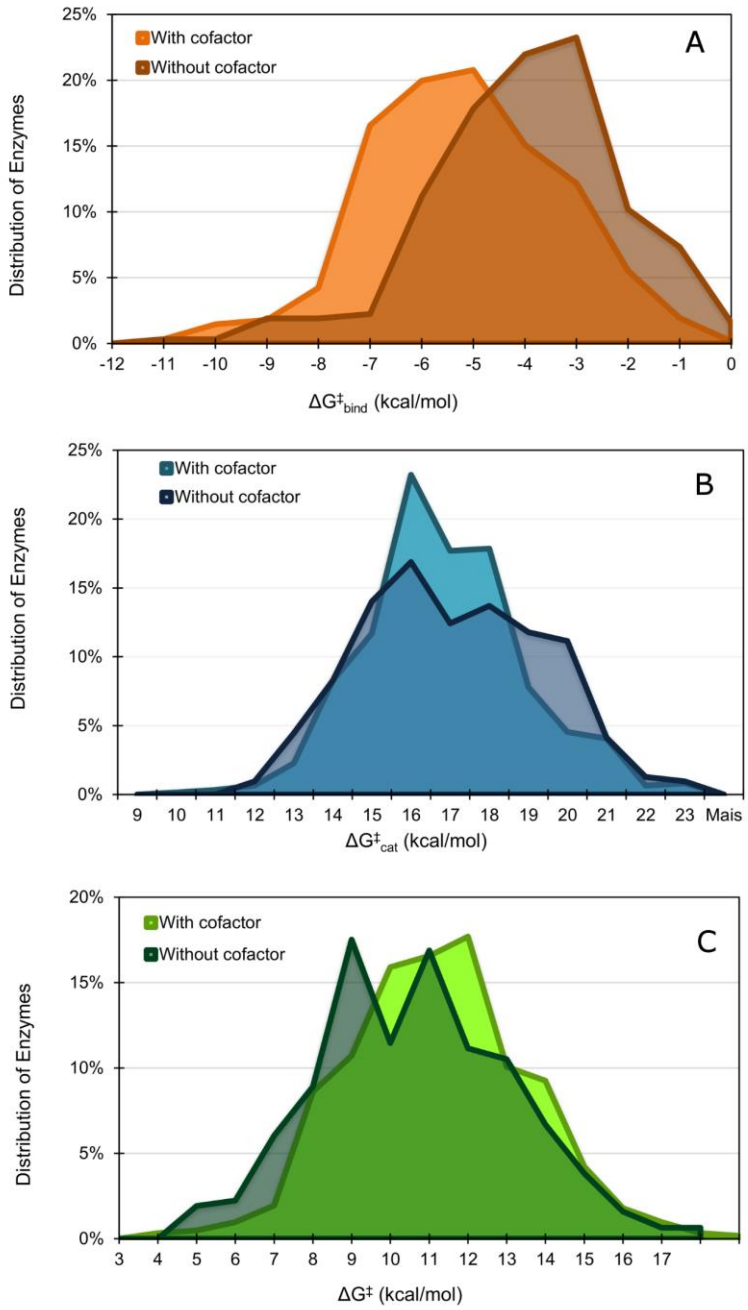


Figure 3.4 Distribution of enzymes with and without cofactors vs. $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$. Panel A represents the distribution of enzymes, with and without cofactors for ΔG_{bind}. Panel B represents that for $\Delta G^{\ddagger}_{cat}$, while Panel C addresses the enzyme efficiency $\Delta G^{\ddagger}$. These results show that in between the enzymes that bind substrates tightly, those that employ cofactors are prevalent.

Table 3.4 shows the distribution of $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$ for the enzymes with and without cofactors. Regarding the binding free energy, Panel A, **Figure 3.4,** shows that enzymes with cofactors have their maximum near -5 kcal/mol, while the distribution of enzymes without cofactors shows that they have a peak at -4 kcal/mol. Panel B (**Figure 3.4**) shows that enzymes that employ cofactors are distributed in a narrow range of $\Delta G^{\ddagger}_{cat}$ ($\Delta G^{\ddagger}_{cat}$ between 15 and 18 kcal/mol) when compared to enzymes without cofactors, which are

distributed in a broader range ($\Delta G^{\ddagger}_{cat}$ between 14 and 21 kcal/mol). The presence of cofactors does not seem to influence the efficiency of enzymes (Panel C – **Figure 3.4**).

3.4.4 Dependence of enzyme parameters on the type of cofactor

There are two types of cofactors employed by enzymes - organic and inorganic. Inorganic cofactors are metals or generic divalent ions and organic cofactors are small organic or metalloorganic molecules.

The average values of ΔG_{bind}, $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}$, for entries that employ different cofactors, are shown in **Table 3.4** that Mg^{2+} is the most common cofactor (136 entries) and it is interesting to see that enzymes that employ this cofactor have, on average, the most negative ΔG_{bind} (-6.3 kcal/mol). This result indicates that probably the presence of this cofactor improves ligand binding. In contrast, enzymes that use this metal seem to have a higher activation free energy (17.1 kcal/mol). The presence of Zn^{2+} is also very common between all enzymes (98 entries). In this case, the ΔG_{bind} is less negative than that of Mg^{2+} (-5.8 vs -6.3 kcal/mol, 97% confidence level), being the $\Delta G^{\ddagger}_{cat}$ smaller (16.4 kcal/mol vs 17.1 kcal/mol, 98% confidence level).

Between enzymes that use metals (only groups with more than 20 entries were considered) the more efficient group is the one that uses Ca^{2+} as cofactor (10.3 kcal/mol), having also the smaller average $\Delta G^{\ddagger}_{cat}$ (16.1 kcal/mol). Despite this, the average value of $\Delta G^{\ddagger}$ from enzymes that employ Ca^{2+} is only statistically different from the one with enzymes that employ Mn^{2+} (96% of confidence levelºlkjhgfgg).

Between groups that depend on coenzymes, FAD is the most common cofactor (83 entries). Enzymes that use this molecule have a small $\Delta G^{\ddagger}_{cat}$ (15.8 kcal/mol), similar to other coenzymes as NAD, or NADP (16.0 and 15.7 kcal/mol, respectively). Enzymes that employ PLP (Piridoxal-5-Phosphate) or ATP have higher values for $\Delta G^{\ddagger}_{cat}$ (16.6 and 17.4 kcal/mol, respectively, being statistically different from the ones that employs $NADP^{+}$, FAD, Heme, ATP and PLP with more than 95% of confidence level). Enzymes dependent on PLP cofactor have the less negative ΔG_{bind} (statistically significant when compared to other enzymes, ~100% confidence level). Enzymes that are dependent on heme prosthetic groups (hemoproteins) seem to be the more efficient between enzymes dependent on cofactors ($\Delta G^{\ddagger}$ = 9.2 kcal/mol), being statistically different (more than 95% of confidence levels) from all of them (except from those that employ NADP+).

Table 3.4 Distribution of enzymes by cofactor type and respective average values of $\Delta G^{\ddagger}cat$, ΔG_{bind} and $\Delta G^{\ddagger}$ and their standard deviations (all in kcal/mol).

	Cofactor	no. of enzymes	ΔG_{bind}	$\Delta G^{\ddagger}_{cat}$	$\Delta G^{\ddagger}$
Inorganic cofactors	Zn^{2+}	98	-5.8 ± 1.7	16.4 ± 2.1	10.6 ± 2.4
	Mg^{2+}	136	-6.3 ± 1.8	17.1 ± 2.3	10.8 ± 2.2
	Mn^{2+}	38	-5.1 ± 1.9	16.4 ± 1.9	11.3 ± 2.0
	Ca^{2+}	32	-5.8 ± 2.3	16.1 ± 2.2	10.3 ± 2.2
Organic cofactors	NAD^{+}	47	-5.6 ± 2.0	16.0 ± 2.1	10.4 ± 2.2
	$NADP^{+}$	47	-6.0 ± 1.8	15.7 ± 1.9	9.7 ± 2.2
	FAD	83	-5.8 ± 1.8	15.8 ± 1.8	10.0 ± 2.3
	Heme	29	-6.2 ± 1.8	15.4 ± 1.4	9.2 ± 1 .1
	ATP	50	-5.8 ± 1.5	17.4 ± 1.6	11.7 ± 1.9
	Piridoxal-5-Phosphate	37	-4.0 ± 1.2	16.6 ± 2.0	11.3 ± 2.4

There are enzymes that employ more than one cofactor and it is useful to find out whether they are more efficient that those that employ only one cofactor. In **Table 3.5** are shown the average values for $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$ for enzymes that employ a single cofactor and for enzymes that employ more than one cofactor. The results show that the number of cofactors is important for the efficiency of an enzyme, influencing $\Delta G^{\ddagger}_{cat}$ and ΔG_{bind}. Enzymes that use more than one cofactor, seem to be more efficient than enzymes that depend on only one cofactor (10.0 vs 10.9 kcal/mol, ~100% confidence level). For enzymes with more than one cofactor, the average value of $\Delta G^{\ddagger}_{cat}$ is smaller than for enzymes with only one cofactor (15.8 vs 16.4 kcal/mol, ~100% confidence level). The average value of ΔG_{bind} for enzymes with more than one cofactor is also more negative (-5.8 vs -5.6 kcal/mol, 84% confidence level).

Table 3.5 Average $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\triangle G^{\ddagger}$ values (in kcal/mol) for enzymes that employ one cofactor or more.

no. of cofactors	no. of enzymes	ΔG_{bind}	$\Delta G^{\ddagger}_{cat}$	$\Delta G^{\ddagger}$
One	**513**	-5.6 ± 1.9	16.4 ± 2.0	10.9 ± 2.3
more than one	**107**	-5.8 ± 1.6	15.8 ± 1.4	10.0 ± 1.8

Summarizing, all these results show that, in general, the presence of cofactors is associated to an increase in the efficiency of enzymes, decreasing the ΔG_{bind}, $\Delta G^{\ddagger}_{cat}$ or both. In the analysis by cofactor identity, the results show that, in general, cofactors that have lower $\Delta G^{\ddagger}_{cat}$, have a slightly higher ΔG_{bind}, with the opposite also true (higher $\Delta G^{\ddagger}_{cat}$,

lower ΔG_{bind}). The number of cofactors seems to be important also, with the presence of more than one cofactor important to increase enzyme efficiency.

3.4.5 Dependence of enzymes parameters on the number of polypeptide chains

When classified based on the number of polypeptide chains, enzymes can be divided in two types: monomeric or oligomeric. Monomeric enzymes are composed by only one subunit (monomer) and oligomeric enzymes are composed of multiple polypeptide chains. Does the number of polypeptide chains affect the efficiency of the enzyme? Does it influence the activation barrier or the substrate binding energy? To answer these questions, we compared the average and corresponding standard deviations of $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$ in both monomeric and oligomeric enzymes. The number of entries for dimers and tetramers is large enough to apply statistics, however trimers, pentamers, hexamers and octamers do not exist in sufficient number to carry out a similar study. Taking this into account, the enzymes were divided in two groups only: monomers or oligomers, and the results are shown in **Table 3.6**. Oligomeric enzymes have been divided in homo-oligomeric or hetero-oligomeric, to determine if the fact of having equal or different subunits influences the studied parameters.

Table 3.6 Average $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$ for monomeric and oligomeric enzymes. Oligomeric enzymes have been divided in homo or hetero-oligomeric enzymes. The results are presented in kcal/mol.

Oligomerization	no of enzymes	ΔG_{bind}	$\Delta G^{\ddagger}_{cat}$	$\Delta G^{\ddagger}$
Monomer	181	-5.9 ± 1.9	16.6 ± 2.1	10.7 ± 2.2
Oligomer	543	-5.5 ± 1.8	16.5 ± 2.1	11.0 ± 2.4
Homo-oligomer	434	-5.5 ± 1.8	16.3 ± 2.1	10.9 ± 2.4
Hetero-oligomer	66	-6.1 ± 2.0	16.7 ± 2.2	10.7 ± 2.1

The results suggest that oligomerization decreases the efficiency of enzymes (11.0 vs 10.7 kcal/mol, 83% confidence level). Furthermore, a higher $\Delta G^{\ddagger}$ is influenced by a less negative $\Delta G_{bind,}$ which means that oligomeric enzymes do not bind their substrates as well as monomeric enzymes (-5.5 vs -5.9 kcal/mol, 98% confidence level). The values for $\Delta G^{\ddagger}_{cat}$ are statistically similar between these two types of enzymes16.6 kcal/mol and 16.5 kcal/mol for monomeric and oligomeric enzymes, respectively).

To verify if the results are similar in each class of enzymes, the same analysis has been performed per class and the results are shown in **Table 3.7**.

There are more **oxidoreductases** (E.C 1) organized as oligomers than as monomers (126 vs 36). This could mean that this class of enzymes has evolved to be organized as oligomers to improve their efficiency. However, disappointingly, when we compare the monomeric and oligomeric enzymes of this class the analyzed parameters seem to be very similar between them.

Table 3.7 Average $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$ for monomeric and oligomeric enzymes divided by enzyme class. The results are shown in kcal/mol.

Class	Oligomerization	no of enzymes	ΔG_{bind}	$\Delta G^{\ddagger}_{cat}$	$\Delta G^{\ddagger}$
EC.1. Oxidoreductases	Monomer	36	-5.7 ± 1.9	16.0 ± 1.6	10.3± 2.0
	Oligomer	126	-5.6 ± 1.8	15.9 ± 1.6	10.2 ± 1.8
EC.2. Transferases	Monomer	33	-5.9 ± 1.5	18.4 ± 2.3	12.5± 2.2
	Oligomer	86	-5.8 ± 2.1	16.9 ± 2.0	11.2 ± 2.3
EC.3. Hydrolases	Monomer	90	-5.8± 2.1	16.1 ± 1.9	10.3 ± 2.1
	Oligomer	182	-5.5 ± 2.0	17.0 ± 2.3	11.4 ± 2.6
EC.4. Lyases	Monomer	11	-7.4 ± 1.7	16.7 ± 2.5	9.2 ± 2.2
	Oligomer	59	-5.2 ± 1.6	16.4 ± 1.8	11.2 ± 2.3
EC.5. Isomerases	Monomer	5	-5.1 ± 0.9	16.4 ± 1.6	11.3 ± 2.0
	Oligomer	55	-4.9 ± 1.2	14.9 ± 2.4	11.6 ± 1.2
EC.6. Ligases	Monomer	6	-6.4 ± 1.1	17.1 ± 0.5	10.7 ± 1.2
	Oligomer	35	-5.5 ± 1.6	17.2 ± 1.4	11.7 ± 2.0

As far as **transferases** are concerned, the results suggest that there are more oligomers than monomers (86 vs 33) and oligomerization enhances the catalytic efficiency (the average value of $\Delta G^{\ddagger}$ is 11.2 kcal/mol vs. 12.5 kcal/mol observed for monomeric enzymes, 99% confidence level). The results also point out that, in this class, oligomeric enzymes have a lower barrier when compared with the monomeric ones (16.9 vs 18.4 kcal/mol, 99% confidence level), with ΔG_{bind} values very similar between them. Overall, considering the results presented, transferases seem to have evolved to be more efficient as oligomers than as monomers.

Within **hydrolases** (E.C.3), oligomerization seems to increase the barrier of the reaction (17.0 vs 16.1 kcal/mol, ~100% confidence level), influencing also the efficiency of these enzymes in the same way (11.4 vs 10.3 kcal/mol, ~100% confidence level).

In the next three classes the number of entries is more reduced. In **lyases**, out of 59 entries, only 11 correspond to monomeric enzymes and therefore we were reluctant in applying statistics to such a small number. However, the results seem to point out that

monomeric lyases bind their substrates better than oligomeric lyases, as observed in previous classes (G_{bind} = -7.4 vs -5.2 kcal/mol). In terms of efficiency, monomers have a lower $\Delta G^{‡}$ (9.2 vs 11.2 kcal/mol) indicative that they are on average more efficient than oligomers. However, it is important to remember that the number of monomeric lyases is very small and these observations may be influenced by the limited number of data available.

In E.C 5 (**isomerases**), the number of monomeric enzymes is equally very small (only 5) and therefore any statistical analysis performed would be meaningless. The same applies to **ligases**, with only 6 monomers.

Summarizing, the overall results show that oligomerization decreases the $\Delta G^{‡}_{cat}$ for most classes of enzymes, with hydrolases as the exception. It is also noticeable that monomeric enzymes on average bind substrates strongly.

As mentioned previously, an additional analysis was performed to understand if there are differences between hetero- and homo-oligomers (**Table 3.6**). The results show that when all chains are similar (homo-oligomeric enzymes) the values for $\Delta G^{‡}_{cat}$ are, on average, lower than when oligomeric chains are different (16.3 vs 16.7 kcal/mol, 82.9% confidence level). On the other hand, the values for ΔG_{bind} are more affected by the type of oligomers that constitute the enzyme complex. Hetero-oligomers seem to facilitate substrate binding (-6.1 vs -5.5 kcal/mol, 97.3% confidence level). As the efficiency is influenced by $\Delta G^{‡}_{cat}$, as well as by ΔG_{bind}, albeit differently, the average values of $\Delta G^{‡}$ are similar between homo- and hetero-oligomers (10.9 vs 10.7 kcal/mol).

3.4.6 Dependence of enzyme parameters on the size of monomers

The size of each monomer in monomeric enzymes can be significantly different. In all entries that we have collected, monomers' sizes range between 150 and 1306 amino acid residues. Does the size of monomers influence enzyme's efficiency? To evaluate how $\Delta G_{bind,}$ $\Delta G^{‡}_{cat}$ and $\Delta G^{‡}$ change with enzyme's size, we divided them in three different groups, selected by size (small, medium and large enzymes), based on the same division performed in our previous work [164]. Small enzymes have less than 274 amino acid residues, medium enzymes have between 275 and 499 amino acid residues and large enzymes have more than 500 amino acid residues.

The results show that increasing the size of monomers increases the efficiency of enzymes ($\Delta G^{‡}$= 11.0 kcal/mol for small enzymes, 10.7 kcal/mol for medium enzymes and 10.3 kcal/mol for large enzymes) (**Table 3.8**). This progressive increment on the catalytic efficiency with monomer size happens because large enzymes seem to lower the

activation free energy of catalyzed reactions ($\Delta G^{\ddagger}_{cat}$ = 17.3 kcal/mol for small enzymes, 16.4 kcal/mol for medium size enzymes and 16.0 kcal/mol for larger enzymes). On the other hand, the results also point out that increasing the size of monomers decreases substrate binding (ΔG_{bind} = -6.3 kcal/mol for small enzymes, -5.9 kcal/mol for medium enzymes and -5.7 for larger enzymes).

The analysis of the monomers by class would be interesting, however it would not be statistically correct to perform it, because there are only a few examples of some classes in each of the groups that we have defined.

Table 3.8 Average $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$ for small, medium and large enzymes. Small enzymes have less than 274 amino acid residues, medium enzymes have between 275 and 500 amino acid residues and large enzymes have more than 500 amino acid residues. All values are shown in kcal/mol.

	monomer's size	no. of enzymes	$\Delta G_{binding}$	$\Delta G^{\ddagger}_{cat}$	$\Delta G^{\ddagger}$
All enzymes	**small (<275 aa)**	39	-6.3 ± 1.5	17.3 ± 2.3	11.0 ± 2.0
	medium (275-500 aa)	59	-5.9 ± 2.2	16.3± 2.1	10.4 ± 1.9
	large (>500 aa)	80	-5.7 ± 1.8	16.0 ± 1.7	10.3 ± 2.0

3.4.7 Dependence of enzyme parameters on cell location

Different enzymes exist in different locations within the cell. They also exist in the extracellular medium. The intracellular environment contains many proteins as well as other molecular compounds that influence the properties of enzymes. Moreover, the quantity of available water is reduced within the cell when compared with the extracellular environment. This is rather important as some enzymes, such as hydrolases, need water molecules to undergo their chemical reactions. Taking this into account, we expected a dependence of enzyme parameters on their location and, accordingly, we separated the enzymes into two groups: intracellular and extracellular. Subsequently, we computed the mean values of $\Delta G_{bind,}$ $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}$, and their corresponding standard deviations, for extracellular as well as intracellular enzymes. The results are shown in **Table 3.9**.

Table 3.9 Average $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$ for extracellular and intracellular enzymes. All values are represented in kcal/mol.

Location	no. of enzymes	ΔG_{bind}	$\Delta G^{\ddagger}_{cat}$	$\Delta G^{\ddagger}$
extracellular	68	-5.4 ± 2.2	15.6 ± 1.4	10.1 ± 2.2
intracellular	631	-5.7± 1.7	16.5 ± 2.0	10.7 ± 2.2

The results show that the efficiency of extracellular enzymes is higher compared to intracellular enzymes (10.1 vs 10.7 kcal/mol, 96.4% confidence level). Extracellular enzymes have a significantly lower activation free energy (15.6 vs 16.5 kcal/mol, ~100% confidence level), although they have a slightly less negative binding free energy (-5.4 vs -5.7 kcal/mol, 73% confidence level).

The number of intracellular enzymes is significantly higher than extracellular enzymes, as expected, since most of the biological reactions occur inside the cells. These enzymes are responsible for different reactions in different cell regions such as membrane or different organelles (mitochondria, chloroplast, lysosome and others). To understand if cell location is related with enzyme parameters we proceeded by analyzing $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$ for intracellular enzymes acting in different cell compartments.

For this analysis, we only considered groups that contained at least more than 20 entries.

Table 3.10 shows the results for enzymes localized in the cytoplasm, nucleus, membrane, mitochondrion, periplasm and peroxisome.

Most of the intracellular enzymes act on the cytoplast (317), having an average $\Delta G^{\ddagger}_{cat}$ of 16.9 kcal/mol, an average ΔG_{bind} of -5.7 kcal/mol and an efficiency of 11.2 kcal/mol. This is expectable as the cytoplasm corresponds to the major part of the volume of the cell. These values are not the lowest ones, but they agree with the average values for all enzymes (discussed in the first topic of the results).

The most efficient enzymes seem to be those that act on the periplasm. The results show that these enzymes have the smaller activation free energy ($\Delta G^{\ddagger}_{cat}$ = 15.2 kcal/mol), which makes them the most efficient between the selected groups. The periplasm is a compartment characteristic of Gram-negative bacteria located between the inner and outer membranes that contains many proteins performing different functions. Many of the enzymes located in this region are responsible to cleave/change large molecules that can be transported more easily to the bacteria cytoplasm.

Nuclear enzymes have the larger $\Delta G^{\ddagger}_{cat}$ (17.4 kcal/mol) and are the less efficient (ΔG_{bind} = 11.4 kcal/mol). These values indicate that the reactions catalyzed by these enzymes are slower when compared to those catalyzed by enzymes located in other cellular compartments. On the other hand, these enzymes facilitate substrate binding, which can be shown by the more negative value of ΔG_{bind} (-6.0 kcal/mol).

Table 3.10 Average $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$ for enzymes in different cell locations. The results are shown in kcal/mol.

Cell location	no. of enzymes	ΔG_{bind}	$\Delta G^{\ddagger}_{cat}$	$\Delta G^{\ddagger}$
Periplasma	28	-5.9 ± 1.3	15.2 ± 1.4	9.3 ± 2.6
Mitochondrion	71	-6.2 ± 2.0	16.0 ± 1.9	9.8 ± 2.3
Peroxisome	23	-4.6 ± 1.9	15.4 ± 1.7	10.8 ± 2.1
Membrane	65	-5.6 ± 1.8	16.4 ± 2.2	10.8 ± 2.5
Cytoplasm	317	-5.7 ± 1.6	16.9 ± 2.0	11.2 ± 2.1
Nucleus	41	-6.0 ± 2.1	17.4 ± 1.9	11.4 ± 2.1

Mitochondria is the organelle responsible for the generation of metabolic energy in eukaryotic cells. The energy that is derived from the break of carbohydrates and fatty acids is converted to ATP in this organelle. It contains enzymes encoded by their own genome and imported from the cytosol. Our results show that mitochondrion enzymes have, on average, the lowest binding free energy, being also some of the most efficient ones.

Peroxisome are organelles that also contain many enzymes involved in different reactions related with energy metabolism. A large number of substrates are broken down by oxidative reactions in these organelles. Analyzing the results obtained in our work, we notice that enzymes acting on peroxisomes are also very efficient, with this efficiency more influenced by a small $\Delta G^{\ddagger}_{cat}$ (15.4 kcal/mol).

It is interesting to see that there is an increase on average enzyme efficiency from nucleus to membrane, passing through cytoplasm, and ending in the periplasm (in Gram-negative bacteria).

3.4.8 Dependence of the enzyme parameters on the temperature

Different enzymes have different temperatures to perform their optimal activity. This temperature is strictly related with environment conditions. To evaluate if temperature is related with the enzyme parameters analyzed in this work, $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$ were computed to two different groups of enzymes: mesophilic enzymes and thermophilic enzymes, a type of extremophile enzymes. Mesophilic enzymes have their optimal activity at temperatures that range between 20 ºC and 45ºC, while thermophilic enzymes have optimal temperatures at higher values (>45ºC). The same division were made in our previous work with hydrolases [164].

Table 3.11 Average values of $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$ for mesophilic and thermophilic enzymes. The results are shown in kcal/mol. Confidence level of the two-tailed *t*-tests are presented too.

Type	no of enzymes	ΔG_{bind}	$\Delta G^{\ddagger}_{cat}$	$\Delta G^{\ddagger}$
mesophilic	861	-5.6 ± 1.8	16.3 ± 2.1	10.7 ± 2.3
thermophilic	73	-5.1 ± 2.2	17.7 ± 2.2	12.3 ± 2.3
two-tailed *t*-test confidence level		98%	~100%	~100%

The results show that the efficiency of thermophilic enzymes is significantly lower when compared with mesophilic enzymes (10.7 vs 12.3 kcal/mol, ~100% confidence level). This result is explained because compared to thermophilic enzymes, mesophilic ones bind the correspondent substrate in a better way (-5.6 vs -5.1, 98% confidence level) and have a significantly lower activation free energy (16.3 vs 17.7, ~100% confidence level) (**Table 3.11**). The thermophilic enzymes, present in extremophilic organisms, usually catalyze their reactions in adverse conditions. Accordingly, the observed differences between mesophilic and thermophilic enzymes are expectable. These results, obtained for all entries in our database, are similar to the results obtained for hydrolases in our previous work.

3.4.9 Dependence of enzyme parameters on substrate specificity

Enzymes can act on a single substrate or in different compounds (promiscuous enzymes). To understand if the enzyme specificity is related to more negative binding energies, or to lower activation barriers, the average values of $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$ and respective standard deviations were analyzed dividing the enzymes in two groups: (1) specific enzymes, which have a single specific substrate and (2) promiscuous enzymes, which have more than one substrate,

The results are presented for all entries collected, as well as for each class of enzymes (**Table 3.11**). Looking at the results for all classes, it is visible that specific enzymes are more efficient than promiscuous enzymes (10.6 vs 11.0 kcal/mol, 98% confidence level). This is because specific enzymes bind substrates better than promiscuous enzymes as expected (-5.8 vs -5.5 kcal/mol, 99% confidence level). However, the average value of activation free energy is not really statistically different (16.3 vs 16.4 kcal/mol, 35% confidence level for being different).

Table 3.12 Average $\Delta G^{\ddagger}_{cat}$, ΔG_{bind} and $\Delta G^{\ddagger}$ for specific enzymes (one substrate only) and for promiscuous enzymes (different substrates). The results are grouped by enzyme class and they are shown in kcal/mol.

		no. of enzymes		ΔG_{bind}		$\Delta G^{\ddagger}_{cat}$		$\Delta G^{\ddagger}$	
	no. of substrases	**1**	**>1**	**1**	**>1**	**1**	**>1**	**1**	**>1**
Class	EC.1. Oxidoreductases	110	103	-5.8 ± 1.9	-5.6 ± 1.7	15.7 ± 1.7	15.9 ± 1.6	9.9 ± 2.0	10.2 ± 2.1
	EC.2. Transferases	58	105	-5.9 ± 2.2	-5.7 ± 1.7	16.9 ± 2.0	17.3 ± 2.0	11.0 ± 2.2	11.7 ± 2.2
	EC.3. Hydrolases	60	279	-6.1 ± 1.8	-5.4 ± 1.9	16.7 ± 2.1	16.5 ± 2.2	10.6 ± 2.6	11.1 ± 2.4
	EC.4. Lyases	49	33	-5.6 ± 1.5	-5.5 ± 1.8	16.7 ± 2.0	16.2 ± 1.9	11.0 ± 1.8	10.8 ± 2.9
	EC.5. Isomerases	54	28	-5.0 ± 1.9	-4.9 ± 1.10	15.8 ± 2.3	14.0 ± 2.3	10.7 ± 2.4	9.1 ± 2.9
	EC.6. Ligases	46	6	-6.1 ± 1.4	-3.7 ± 1.10	17.3 ± 2.0	17.5 ± 1.7	11.3 ± 2.4	13.9 ± 2.9
	All enzymes	377	554	-5.8 ± 1.8	-5.5 ± 1.8	16.4 ± 2.0	16.4 ± 2.1	10.6 ± 2.2	11.0 ± 2.4

In the case of oxidoreductases (EC.1), the results suggest that specific enzymes are more efficient than promiscuous enzymes (9.9 vs 10.2 kcal/mol, 77% confidence level). In transferases (EC.2) the same trend is verified (11.0 vs 11.7 kcal/mol, 93% confidence level). In both cases the average binding energies (ΔG_{bind}), as well as the average activation free energies ($\Delta G^{\ddagger}_{cat}$), are lower for specific enzymes, even though the differences between specific and promiscuous enzymes are not statistically significant.

In hydrolases (EC. 3), analyzed in our previous work, substrate-specific entries have also lower $\Delta G^{\ddagger}$ (10.6 vs 11.1 kcal/mol, 89% confidence level), being more catalytically efficient than promiscuous ones. This occurs because they have a tendency to bind strongly to the substrate (ΔG_{bind} = -6.1 vs -5.4, 99% confidence level).

The results for lyases (EC. 4) show that the parameters do not change with substrate specificity.

the results for isomerases (EC. 5) show that the average activation free energy is lower for the promiscuous enzymes of this class (14.0 vs 15.8 kcal/mol, ~100% confidence level). Therefore, this group of enzymes seems to be more efficient than the substrate-specific ones (9.1 vs 10.7 kcal/mol, 98% confidence level).

In turn, in the last class (Ligases – EC. 6) the results show that specific enzymes may be more efficient than promiscuous enzymes (11.2 vs 13.8 kcal/mol), maybe because these enzymes have a more negative energy of binding (-6.1 vs -3.6 kcal/mol). It is important to note that, in this class, the number of promiscuous enzymes is very low (6 entries) and, therefore, a statistical analysis is not particularly meaningful.

2.5 Discussion

Kinetic data was analyzed in detail in this work and related with available structural and biological experimental data, for each class of enzymes. Despite the large number of enzymes that were analyzed and the differences between them, similarities and trends were found. Here we discuss the significant findings that we present in this work as well as some open questions that remain unclear in the origin of the similar catalytic power of the huge diversity of enzymes.

2.5.1 Enzyme classes have similar ΔG_{bind} and $\Delta G^{\ddagger}_{Cat}$

The binding free energies, ΔG_{bind}, for all enzymes range from -12.3 to -0.6 kcal/mol for all 934 entries, with 85% fall in a narrow range between -3 and -7 kcal/mol, having a mean of -5.6 ± 1.8 kcal/mol. Looking at the values of each enzyme class, we notice that they are very similar between them.

The activation free energies, $\Delta G^{\ddagger}_{cat}$, of all entries analyzed, also fall in the range between 8 and 23 kcal/mol, with 94% in the range from 14 to 21, presenting an average value of 16.4 kcal/mol.

Enzymes lower the energy barriers for catalysis, and therefore they lower the apparent activation free energies ($\Delta G^{\ddagger}_{cat}$) and the binding free energies (ΔG_{bind}) to a narrow range of values, for very different reactions, which have very different free energies in solution chemistry[1]. Due to the thin range of the values that these two parameters adopt, as well as to compensation and/or correlations between them, the catalytic efficiency of enzymes is also narrowed down to a very specific range of values.

It is important to note that in our previous work, we arrived at similar conclusions for hydrolases only[164].

2.5.2 Different classes of enzymes, similar efficiency

Different classes of enzymes catalyze different reactions. However, we have observed that the catalytic efficiency falls in a very narrow range of values for all classes ($\Delta G^{\ddagger}$ = 8 to 14 kcal/mol for 87% of all enzymes). This happens because different classes of enzymes have different strategies to lower $\Delta G^{\ddagger}_{cat}$ and ΔG_{bind} to achieve those values. It is interesting to observe that classes of enzymes that tend to improve the catalytic efficiency by lowering the activation free energy have less negative ΔG_{bind} values, and vice-versa. Isomerases (EC.5), which have the lowest average value of $\Delta G^{\ddagger}_{cat}$ among all enzyme classes (15.2 ± 2.4 kcal/mol), have the higher average value of ΔG_{bind} (-5.0 ± 1.8 kcal/mol),

which means that the catalytic power of this class of enzymes is related to lower activation free energies, rather than to substrate binding. In turn, ligases (EC.6) that seem to have the higher activation free energy ($\Delta G^{\ddagger}_{cat}$ = 17.4 ± 1.6 kcal/mol), have the more negative value for ΔG_{bind} (-5.8 ± 1.5 kcal/mol), indicating that this class of enzymes improves their catalytic efficiency by binding their substrates more tightly.

Taking this into account, it is plausible to think that, in order to improve the catalytic efficiency of an enzyme, evolution can resort to improving either the binding of the substrate or the activation barrier of the reaction.

How do enzymes seem to improve substrate binding?

According to the results presented in this work, there seems to be at least three ways in which enzymes improve substrate binding, presenting more negative values of ΔG_{bind}:

specificity for one substrate only rather than multiple substrates. We have verified this fact also in our previous work for hydrolases and it is transferable to all classes of enzymes.

reaching for the presence of Mg^{2+}, as it contributes to tighter binding of the substrate.

by being small, since smaller enzymes also seem to be associated to stronger substrate binding.

How do enzymes lower the activation barrier?

Again, we have noticed some ways, adopted by enzymes, in order to lower their activation barriers:

the presence of more than one cofactor contributes to lower $\Delta G^{\ddagger}_{cat}$.

larger enzymes are related to small values of $\Delta G^{\ddagger}_{cat}$. These results are also in agreement with our previous analysis for hydrolases.

activation free energies tend to be lower for extracellular enzymes or for enzymes that act on periplasm or peroxisomes.

2.6 Conclusions

The study presented here shows that, despite the existence of a large diversity of enzyme structures catalyzing different chemical reactions with distinctive substrates and environmental conditions, all enzymes have activation free energies, substrate binding free energies and catalytic efficiencies that fall in a narrow range of values. A few strategies that enzymes employ to improve their efficiency have been identified. In fact, some enzymes decrease their activation free energies or bind their substrates tightly resorting to cofactors. Others achieve these characteristic values by having different

numbers of chains and different number of amino acid residues. Their location in the cell also affects these properties. **Table 3.13** summarizes some general trends observed for all enzymes, which could be useful for future work.

Table 3.13 Average values of ΔG_{bind} and $\Delta G‡cat$ for all enzyme entries and a systematization of the main conclusions of this work.

Properties	Trends for all enzymes (average values)
ΔG_{bind} (kcal/mol)	-5.6 ± 1.8
$\Delta G^{‡}_{cat}$ (kcal/mol)	16.4 ± 2.1
$\Delta G^{‡}$(kcal/mol)	10.8 ± 2.4
Classes	-Isomerases have the lowest $\Delta G^{‡}_{cat}$ (15.2±2.4 kcal/mol) and the higher ΔG_{bind} (-5.0±2.8 kcal/mol). -Ligases have the higher $\Delta G^{‡}_{cat}$ (17.4±1.6 kcal/mol) and the lowest ΔG_{bind} (-5.8±1.5 kcal/mol). -The overall outcome is very similar for all classes.
Cofactors	-In general enzymes with cofactors lower the $\Delta G^{‡}$ value. -Mg^{2+} facilitates substrate binding. -More than one cofactor lower the $\Delta G^{‡}_{cat}$ values.
Oligomerization	-Monomeric enzymes are more efficient than oligomeric ones.
Size	-Small enzymes (< 250 aa residues) facilitate substrate binding but have higher activation free energies.
Location	-Extracellular enzymes are more efficient than intracellular ones.
Substrate specificity	-Substrate specific enzymes have an average binding free energy lower than promiscuous ones.

This systematization helps to ascertain that all chemistry of life seems to have its own specific time-scale. Enzymes permit to synchronize and fit the chemistry of living beings to this time-scale, independently of the type of reaction that they catalyze, the organism and their structural diversity.

CHAPTER 4. Reaction Mechanism of Human Renin Studied by Quantum Mechanics/Molecular Mechanics (QM/MM) Calculations

Ana Rita Calixto†, Natércia Fernandes Brás, Pedro Alexandrino Fernandes† and Maria João Ramos

†UCIBIO, REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal,

In this chapter and in the next one, the catalytic mechanism of two different enzymes were explored and described by computational methods. Although these two works are not directly related with the origin of the catalytic power of enzymes, they are important to understand the atomistic details of the reaction that both enzymes (renin and PatG macrocyclase - next chapter) catalyze. In the present chapter the catalytic mechanism of human renin was explored, using a wild type and a mutated substrate. It was verified that this enzyme follows a mechanism with three sequential steps, typical from aspartic proteases, with both the wild type and the mutated substrate. A comparison between human renin and mouse renin (published before) was also explored.

Regarding the contributions to the paper, Ana Rita Calixto performed all the calculations and wrote the first draft manuscript, which was revised through contribution of all authors. This work was published in the *ACS Catalysis*, and the following content is almost an integral transcription of the published version.

Calixto, A. R.; Bras, N. F.; Fernandes, P. A. and Ramos, M. J., ***Reaction Mechanism of Human Renin Studied by Quantum Mechanics/Molecular Mechanics (QM/MM) Calculations***. ACS Catalysis, 2014, 4 (11), 3869-3876 **DOI: 10.1021/cs500497f7**

4.1 Abstract

In this paper we present the catalytic mechanism of human renin computationally investigated using an ONIOM QM/MM methodology (B3LYP/6-31G(d):AMBER), with final energies calculated at the M06/6-311++G(2d,2p):AMBER level of theory. It was demonstrated that the full mechanism involves three sequential steps: i) a nucleophilic attack of a water molecule on the carbonyl carbon of the scissile bond, resulting in a very stable tetrahedral gem-diol intermediate, ii) a protonation of the peptidic bond nitrogen and iii) a complete breakage of the scissile bond. The activation energy barrier obtained for the angiotensinogen hydrolysis by renin was calculated as 22.0 kcal.mol^{-1}, which is consistent with the experimental value, albeit slightly above. We have shown also that the cleavage of a mutated substrate (Leu10Phe) occurs in a similar way to that of the wild type substrate. These results provide an understanding of the reaction catalyzed by human renin with atomistic detail. This is of particular importance because this enzyme plays a special role in the control of the Renin-Angiotensin system and consequently it is at the center of current hypertension therapy.

KEYWORDS

Catalytic mechanism, Hypertension, ONIOM, QM/MM, renin

4.2 Introduction

This work explores the catalytic mechanism of human renin with atomic-level detail. Computational methods have been employed for that purpose. Behind the fundamental biochemical knowledge that is gained, the results also facilitate the discovery of new inhibitors with therapeutic potential.

The Renin Angiotensin System (RAS) is the principal regulator of fluid homeostasis and blood pressure. This multi-enzymatic cascade system begins with a two-step hydrolysis of the angiotensinogen peptide. In the first enzymatic reaction, angiotensinogen is cleaved by renin, resulting in the release of the decapeptide angiotensin I (Ang I). Subsequently, the angiotensin converting enzyme (ACE) hydrolyses Ang I, releasing the two terminal residues and generating the octapeptide angiotensin II (Ang II), an agonist of the AT1 receptor that, when activated, induces an increase in blood pressure.

Excessive activation of this system leads to a blood pressure increase and consequently to hypertension, which is nowadays a very important worldwide public health challenge [175-180].

The hydrolysis of angiotensinogen carried out by renin constitutes the rate-limiting step of the RAS cascade [181-184]. The cleavage takes place between two residues, leucine and valine, and releases the 10 N-terminal residues generating Ang I, which is the only substrate for renin [176,177,185,186] that is known to man. As such, renin is a very attractive target to control the blood pressure. Intervention at this level of the RAS cascade is more specific than intervention more downstream, such as at the ACE enzyme or AT_1 receptor, as no other metabolic pathway is affected. The first inhibition of renin available over the counter (Aliskiren) reached the marked only in 2007, even though the research of bioavailable inhibitors of renin has begun many decades ago [187]. Unfortunately, Aliskiren is associated with significant side effects. Therefore, studies on renin are very important to promote future studies on its inhibition [188-190].

4.2.1 Renin - Relation between structure and function

The monomeric protease renin (EC 3.4.23.15), a member of the aspartic protease superfamily [191], has 340 residues (m.w. of 40 kDa). This enzyme consists of two similar β-sheets lobes with dissimilar sequences. As in many other aspartic proteases, between these lobes there is a deep cleft where the active site is located, burying seven substrate residues.

This active site has two essential, coplanar, aspartic acid residues (Asp38 and Asp236), emanating from different lobes, whose conformation is stabilized by a network of hydrogen

bonds. The carboxyl groups are hydrogen bonded to a water molecule that is essential for renin catalytic activity. Its position close to the substrate, polarized by one aspartate residue, is fundamental for the catalytic mechanism of renin [177,190,192-194]. Renin has a flexible flap close to the active site. An identical structure is also common in other aspartic proteases and it allows covering the active site upon substrate binding. Recent Molecular Dynamics (MD) studies on this enzyme showed that this flap oscillates between an open, semi-open and close conformation. This behavior is important to allow the binding of the substrate and an extensive range of different inhibitors [179,195-197].

4.2.2 Mechanism proposed for aspartic proteases

β-secretase, HIV-1 protease, presenilin and other aspartic proteases have had the details of their catalytic mechanisms described by previous QM/MM and MD simulations[64,198]. The catalytic mechanism of renin was also previously proposed as a case of acid-base catalysis, similar to what has been published for other proteases of this family. However, support for the proposal is based just on the similarities between enzyme family members, and the atomistic details of this reaction have not been described until now. One of the consequences of the close position of the two aspartates is that the pKa of one of them is raised and the other one is lowered, originating different protonation states for each of them (one protonated and other deprotonated) [64]. The postulated mechanism assumes that the catalytic water molecule, which is tightly bound in the active site, is activated by the negatively charged aspartic residue that acts as a general base. Consequently, a nucleophilic attack to the scissile peptidic bond is carried out by the pseudo-hydroxide ion generated *in situ*, giving rise to a gem-diol intermediate. Subsequently, the peptidic nitrogen of the scissile bond is protonated by the second catalytic aspartate and simultaneously a gem-diol hydroxyl is deprotonated by the first, resulting in the cleavage of the peptidic bond **(Figure 4.1**) [64,199-201].

4.2.2.1 The consequences of a Leu10Ile mutation at the substrate, during the catalytic mechanism of human renin

The replacement of a leucine by a phenylalanine at the 10th position of the angiotensinogen, which corresponds to the site of the renin cleavage, is related to the emergence of a hypertensive disorder that is common in pregnancy (preeclampsia) [202]. Due to the importance of this link, we chose to study also the details of the corresponding catalytic mechanism of human renin in the presence of the Leu10Ile mutated substrate.

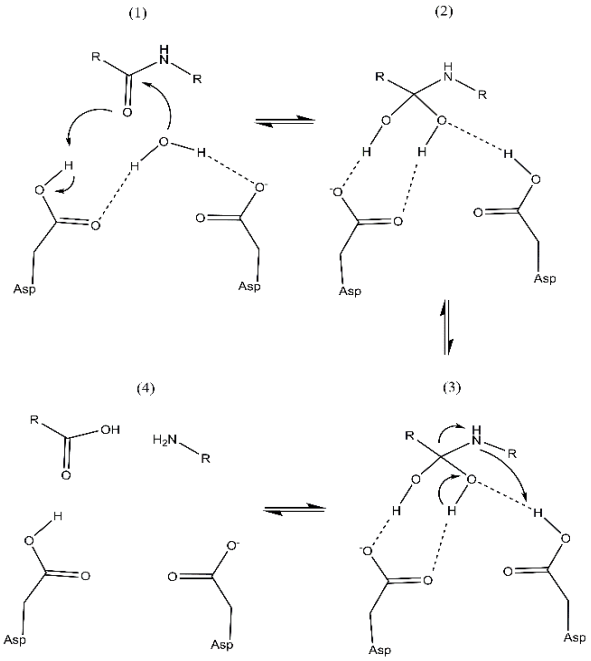


Figure 4.1 General representation of the catalytic mechanism proposed for aspartic proteases. The catalytic water molecule is tightly bound to the catalytic aspartates (1). In the first step, a gem-diol tetrahedral intermediate is formed (2). The breaking of the C-N bond is accompanied by the transfer of a proton to the nitrogen of the leaving amino group (3 and 4).

4.3 Methodology

The X-ray structure of unbound human renin (PDB ID: 2REN, 2.5 Å resolution), was used as starting structure for the present study [191]. The substrate was modeled from a lower resolution structure of renin bound to angiotensinogen (PDB ID: 2X0B, 4.4 Å resolution) [203] and corresponded to the Val3-Asp14 residues. We aligned and superimposed both structures from the PDB 2X0B file and transferred the modeled substrate. In the active site, between the catalytic aspartates (Asp38 and Asp226), a water molecule was added. The software X-leap [204] was used to protonate the complex, assuming physiological protonation states for all residues. According to the proposed mechanism for aspartic proteases [64], one of two catalytic aspartates (Asp38) is protonated at the beginning of the reaction, and therefore we protonated this residue. Eight sodium counter-ions were added with X-leap to compensate for the negative charge of the system. The system was surrounded by a cubic box of TIP3P water molecules having a minimum distance of 12 Å between the protein atoms and the end of box.

A two stages minimization of the geometry using molecular mechanics (parm99 force field) was performed in order to eliminate clashes and bad contacts. The first stage

involved only water molecules and counter ions, keeping fixed the position for the protein atoms. In the second stage all atoms were geometry-optimized. The potential energy surface (PES) for the postulated reaction mechanism was explored at the QM/MM level with the Gaussian 09 software [135]. The reactant state was taken from the previously minimized structure, upon deletion of all water molecules beyond a shell of 5 Å around the enzyme-substrate complex. All counter-ions were far from the active site and, therefore, were removed.

We have employed the ONIOM [205] method as implemented in Gaussian09. The system was partitioned into two layers, a “high layer” treated at the DFT level and a “low layer” treated at the classical molecular mechanics level.

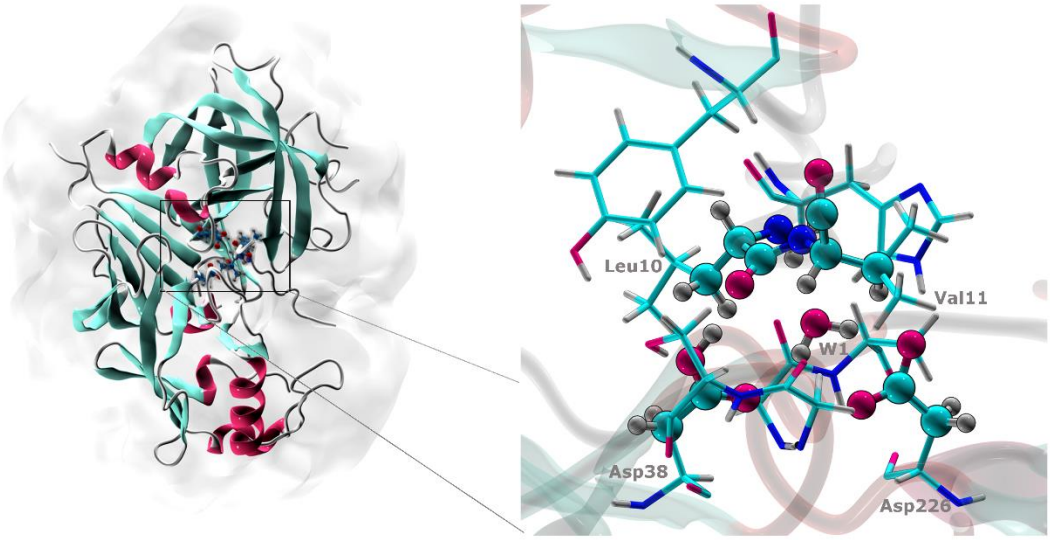


Figure 4.2 QM/MM model used in the calculations. The optimization of the geometries along the reaction coordinate was performed with 33 atoms in the high layer at the B3LYP 6-31G(d) level (represented by balls and sticks). Then the low level layer selected included 138 atoms and was treated with different density functionals (B3LYP, M06, B1B95 and mPWB1K) at the 6-311++G(2d,2p) level of theory (licorice representation).

Different layer divisions and theoretical levels were considered during geometry optimizations and single-point energy calculations. The first were performed with a small high layer (33 atoms) at the B3LYP 6-31G(d) level of theory (**Figure 4.2**), and the low layer with the Amber parm99 force field. The B3LYP functional was employed as it has been shown to provide accurate results for organic molecules [115,206,207].

We used hydrogen as link atoms whenever covalent bonding spanned the QM/MM boundaries, in order to saturate the dangling bonds. The system was further divided into a frozen and a free region. The free region included all the residues that have at least one atom within a 20 Å radius around any atom of the high layer. The remaining atoms, which

were located at the periphery, were kept frozen. The interaction between layers was treated with a mechanical embedding scheme.

Linear scans along the reaction coordinates for each mechanistic step were carried out, starting with 0.10 Å increments to locate the relevant transition states, and subsequently finer 0.05 Å increments were made to refine the transition state geometry. Finally, a full optimization of the transition states was performed, starting from the higher points of the preliminary scans. Atomic point charges for the high layer atoms were determined with the RESP (Restrained Electrostatic Potential) charge fitting program [208,209], using the Merz-Singh-Kollman scheme [210] at each scan step. These PES scans started from a fully optimized structure of the reactants. The intermediates and products were located also through unconstrained geometry optimizations.

Therefore, all stationary points were geometry optimized and the transition state and minima structures were verified by vibrational frequency calculations having exactly one and none imaginary frequencies, respectively. We have calculated also the zero-point energy, the entropy and the thermal corrections for the change from 0 to 298.15 K, which allowed us to obtain the Gibbs energies at physiological temperature. All these calculations were carried out at the B3LYP/6-31G(d):AMBER level of theory. However, to compare the barriers between the wild-type and the Leu10Phe mutant, we used the PES profiles as approximations, which is accurate enough for comparison purposes. A larger high layer was then selected (a set of residues that participate in the reaction or have important interactions with the active site and the substrate), including 138 atoms **(Figure 4.2).** Single point energy calculations were made subsequently, resorting to the electronic embedding scheme and different density functionals (M06, B1B95 and mPWB1W), whose performance for thermodynamics and kinetics is known to be excellent, together with the larger 6-311++G(2d,2p) basis set.

The energies presented in this work are affected by the used methods and the errors associated with them. From experience with other studies conducted in our group, the errors associated with these methods can influence results in approximately 2-4 kcal.mol^{-1} [71,211,212]. These errors may be grounded on several causes such as: the functional chosen, a finite basis set, a high layer with a finite number of atoms, errors associated with molecular mechanics, the use of a single X-ray structure (that corresponds to an average of all molecules in the crystal), and also a truncated substrate instead of the natural substrate.

4.4 Results and Discussion

As all aspartic proteases, renin hydrolyses peptide bonds by an acid/base mechanism that is supposed to start with the formation of a gem-diol intermediate and to end with the cleavage of the peptide bond. In order to understand the cleavage of the bond between positions 10 and 11 of the decapeptide that mimics angiotensinogen, we followed this mechanism with QM/MM calculations.

4.4.1 Hydrolysis of the wild type substrate

4.4.1.1 The structure of the reactants

In the structure of the reactants, the active center water molecule was hydrogen bonded to both catalytic aspartate carboxylates, a conformation adequate for catalysis. Another active site conformation had the two aspartates hydrogen bonded to each other but one of the two was protonated. The latter was more stable, but it was not catalytic productive. The difference between the two conformations was small and transition between both could be seen easily in molecular dynamics simulations. Therefore, in the present work, we started from the first productive conformation, and we used the twelve terminal residues that bind the enzyme cleft as a model of the angiotensinogen substrate [64]. At the beginning of the reaction, the carbonyl oxygen of Leu10 (which belonged to the peptide that was to be cleaved), was hydrogen bonded (1.61 Å) to the carboxylic proton of Asp38 (**Figure 4.3**). A hydrogen bond was also formed between the hydroxyl group of Ser41 and the carboxyl group of Asp38, enhancing the acidity of the aspartate. This serine residue was part of a network of hydrogen bonds, involving Tyr83 and Trp45, which stabilizes the active center conformation. The initial conformation of the reactive aspartates was also stabilized by hydrogen bonding to two glycine residues (Gly39 and Gly220).

4.4.1.2 The first reaction step - The nucleophilic attack of the catalytic water molecule

The catalytic cycle was supposed to begin with the formation of a gem-diol intermediate. As such, the distance between the water oxygen and the carbonyl carbon of the scissile bond was adopted as reaction coordinate. In the optimized reactants, the catalytic water molecule was trapped between Asp38 and Asp226, establishing hydrogen bonds with bond residues with lengths: H_{W1}-O_{Asp226} – 2.00 Å and H_{W1}-O_{Asp38} – 2.16 Å. Such

interactions, and the protonation states of the catalytic aspartates, provided the correct position of the water molecule to promote the nucleophilic attack to the carbonyl carbon.
After partial deprotonation by Asp 226, the resulting hydroxyl anion carried out a nucleophilic attack on the carbonyl carbon of Leu10. Concertedly, Asp38 protonated the Leu10 carbonyl and the Asp226 was protonated by one of the water molecule protons. Full optimization of the transition-state structure and vibrational frequencies analysis revealed just one imaginary frequency (172i cm^{-1}) that coincided with the adopted reaction coordinate (O_{W1} - $C_{carbonyl}$ stretching). The distance between O_{W1} and $C_{carbonyl}$ decreased from 3.20 Å to 1.77 Å, when moving from the reactants (R) to TS1, both illustrated in **Figure 4.3**. The proton of Asp38 was not fully transferred to the carbonyl oxygen at this stage. In the products of this step, the carboxylate group got protonated and a new covalent bond between the water oxygen and the carbonyl carbon of Leu10 was established. The carbonyl carbon changed from sp_2 to sp_3, and the peptide bond became tetrahedral. As can be seen in **Figure 4.3**, this intermediate (INT1) had two hydroxyl groups covalently bonded to the carbon of the scissile bond (gem-diol intermediate). Asp38 and Asp226 both changed their protonation states, and the two hydroxyl groups from the substrate established hydrogen bonds with Asp38 (1.65 Å and 1.64 Å).
Previous studies have provided indications that the aspartic proteases reactions start by a nucleophilic attack of a water molecule on the scissile bond of the substrate, as we tested in this work. However, the nature of the first intermediate is still a matter of controversy. Some studies indicate that the reaction intermediate resulting from the first step is a neutral gem-diol; other studies suggest that a charged hydroxyl anion is stable in the active site of the proteases. In the present study we performed a scan in which the water oxygen was pushed against the carbonyl carbon, and the forming oxyanion was spontaneously protonated by Asp38. We did not find a stationary point with the gem-diol ionized and the Asp38 protonated. This result is in agreement with previous studies, which suggest that the neutral gem-diol intermediate is more stable than a charged oxyanion [213].
The structure of the first intermediate was similar to that described for other members of the aspartic protease family. However, with regards to the TS1 geometry, it was different from that described for other family members, such as BACE-1 or HIV-1 proteases, which involve the formation of a hydroxyl anion at the TS [198]. The other residues around the active center preserved their overall structure along this reaction step. For example, the hydrogen bond network established between Asp38-Ser41-Tyr83-Trp45 was kept intact. However, due to the ionization of Asp38, the hydrogen bond between Ser41 and Asp38 got shortened and the first stabilized the carboxylate group of the latter.

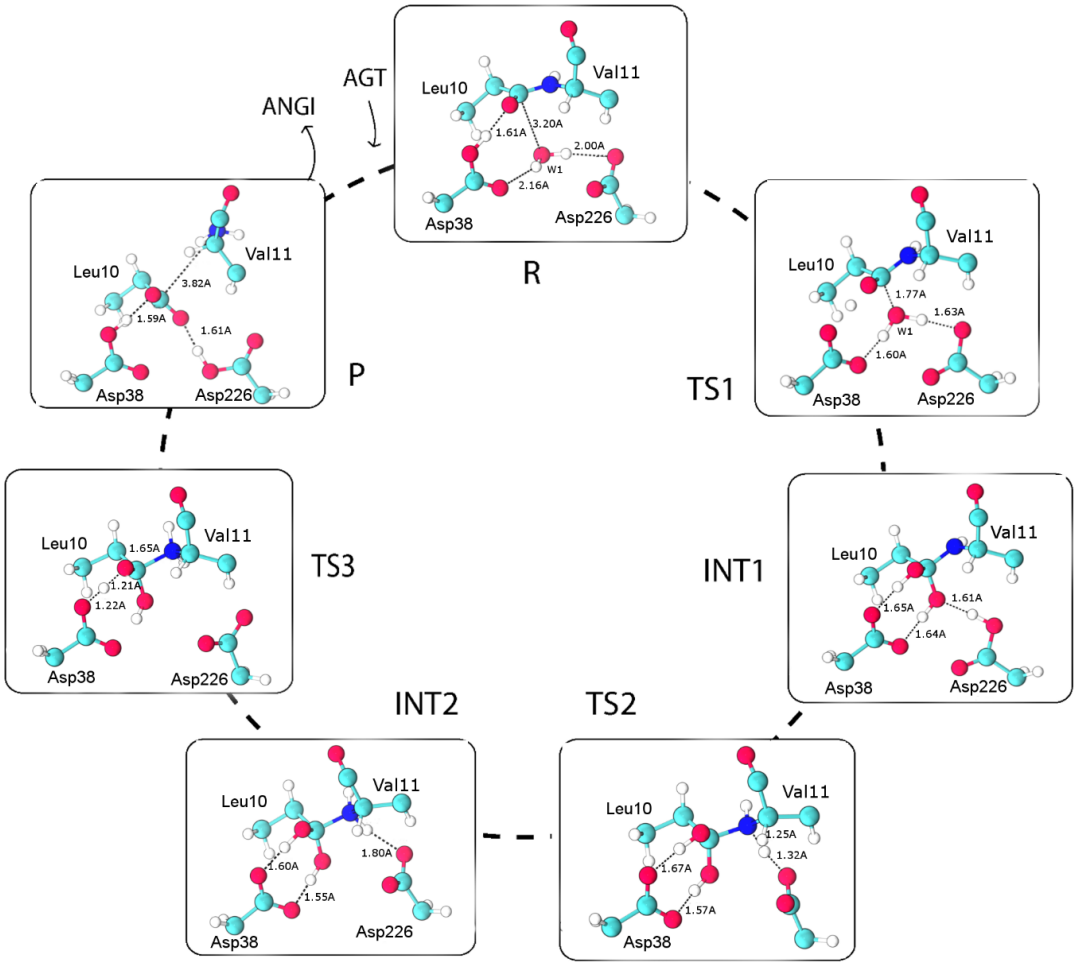


Figure 4.3 Structures of the reactants, intermediates, transition states and products for the cleavage of the Leu10-Val11 peptide bond of angiotensinogen by human renin. Only the QM-treated atoms are represented in the figure for clarity.

The activation free energy for this step was calculated in 22.0.kcal.mol^{-1} with a large high layer (138 atoms) at the M06/6-311++G(2d,2p):AMBER level of theory (**Figure 4.4**). This value is comparable, albeit higher, to the experimental one obtained for the angiotensinogen hydrolysis by renin (16.5 kcal.mol^{-1}) [214].

4.4.1.3 The second reaction step

The mechanism of reaction for aspartic proteases (HIV-1, β-secretase and presenilin) was proposed to occur in two steps: first the formation of a gem-diol intermediate, and second the breakage of the peptide bond of the substrate [64,198]. These studies showed that the hybridization of the scissile bond carbon (sp_3) placed the amine nitrogen in an orientation

that facilitated the attack of the Asp226 carboxyl hydrogen and, consequently, the breakage of the peptide bond and the return of a proton from the gem-diol intermediate to Asp226. In order to describe the second step of renin's catalytic mechanism and taking into account previous studies on aspartic proteases (particularly in mouse renin) we started the second step using the geometry of the INT1. At this stage the residues, such as Asp226, were pre-organized to facilitate the protonation of the nitrogen of the scissile bond. The Asp226 carboxyl proton–peptidic nitrogen distance was taken as reaction coordinate. During this scan we obtained the geometry of the second transition state and intermediate (TS2 and INT2), which are represented in **Figure 4.3**. We optimized the structure of the TS2 and the nature of the transition state was verified by analyzing the vibrational frequencies. We found an imaginary frequency (911i cm^{-1}) that corresponded to the stretching of the atoms involved in the adopted reaction coordinate. The transfer of the proton was associated to a free energy barrier of 13.9 $kcal.mol^{-1}$ (**Figure 4.4**) and took place concomitantly with an increase in the length of the scissile bond.

Some relevant differences between the INT1 and TS2 structures could be seen easily. When the Asp226 acidic proton came closer to the Val11 nitrogen, it broke its hydrogen bond with the gem-diol hydroxyl and changed its rotamer. The Asp226 proton was shared between donor and acceptor at TS2 (distance of 1.25 Å to Leu10 nitrogen and 1.32 Å to the oxygen). The complete transfer of this proton took place only at INT2.

During this second step, the peptidic bond stretched to 1.53 Å at TS2, and to 1.57 Å at INT2 (**Figure 4.3**), meaning that the bond did not completely break during this step. In previous studies of other aspartic proteases at the end of the second reaction step, the scissile bond of the substrate had got broken already. Consistently, the deprotonation of one of the gem-diol hydroxyl groups, back to Asp38, was not observed either. Both catalytic aspartates became negatively charged, while the nitrogen of Leu10 became positive. Previous studies on other aspartic proteases (β-secretase) indicate a rearrangement of the active residues during this step, and a disruption of a hydrogen bond between Ser41 and Tyr83, subsequently to the cleavage of the scissile bond. However, neither the cleavage of the substrate nor this reorganization was observed here, and the network of hydrogen bonds, which was referred to beforehand, between Asp38-Ser41-Tyr83-Trp45, was preserved during this second step [198,215]. A third reaction step was needed to generate the products.

4.4.1.4 The third reaction step

To generate the products, we carried out a linear transit scan along the distance between each gem-diol proton and the carboxyl oxygen to which it was hydrogen bonded. The

structures of the transition state (TS3) and reaction products (P) were easily obtained and are shown in **Figure 4.3.** It was difficult to optimize TS3, and we could not achieve standard convergence criteria. Anyway, its nature was verified by the existence of a single imaginary frequency (596i cm^{-1}). During the progressive transfer of the gem-diol proton to Asp38, the scissile bond of angiotensinogen elongated and spontaneously broke. In the products, the catalytic aspartates were both protonated and two independent peptides were formed, as expected. This reaction step, which led to the angiotensinogen cleavage, had an activation free energy of 16.5 kcal.mol^{-1} (**Figure 4.4).**

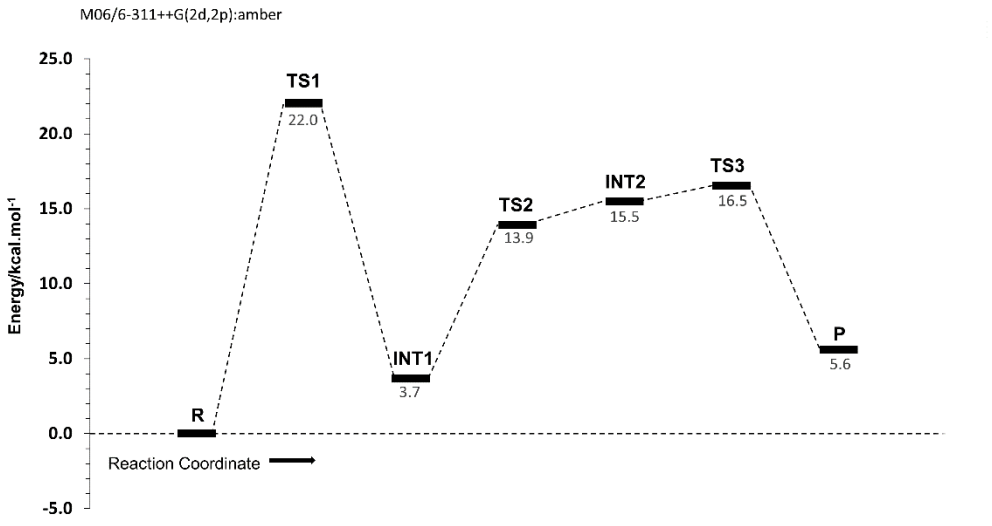


Figure 4.4 Energetic pathway for the hydrolysis reaction of angiotensinogen catalyzed by human renin at the M06/6-311++G(2d,2p):AMBER level with 138 atoms in the high layer.

4.4.2 Hydrolysis of the mutated substrate

Previous studies have linked a specific mutation in angiotensionogen with preeclampsia, a common hypertensive disorder of pregnancy. This mutation, at the site of renin cleavage, corresponds to a replacement of a leucine at the 10^{th} position of angiotensinogen for a phenylalanine [202]. Because of this association, computational methods were used also to describe the hydrolysis of this mutated substrate by renin.

4.4.4.1 The first reaction step – formation of the gem-diol intermediate

The optimized reactants for the mutated substrate were very similar to the wild type substrate described beforehand. The comparison between the complex renin:wild type

substrate and renin:mutated substrate revealed few differences between the two structures. In fact, the only one lied in the substrate in which a Leu residue was replaced by a Phe residue, at position 10. In the optimized reactants, the catalytic water molecule interacted with the Asp dyad through hydrogen bonds. The same strategy that was used for the previously described mechanism, was adopted, and the same reaction coordinates were followed through linear transit scans. In the first reaction step, the distance between the oxygen of the catalytic water molecule and the carbonyl carbon of Phe10 was used as reaction coordinate. The catalytic water molecule performed a nucleophilic attack on the carbonyl group of the scissile bond concerted with protonation of the carbonyl oxygen by Asp38. The structure of the TS1 was identified and has shown that the distance between the water oxygen and the carbonyl carbon decreased from 3.24 Å to 1.70 Å. There were no significant changes in the residues around the active site, as had already happened in the first described mechanism. Overall, this step was very similar to the corresponding step of the wild-type substrate. In the first transition state, the oxygen of the water molecule and the carbonyl group were slightly closer than in the previous mechanism. The Ser41-Tyr83 hydrogen bond was shorter as well. The larger size of phenylalanine, compared to leucine, may explain these differences. Despite these aspects, the first reaction step was very similar overall to what was described for the reaction with the wild type substrate. As in the previous mechanism the protonation of the carbonyl group was completely accomplished at the end of this first step. In the end a gem-diol intermediate as obtained, with the hybridization of the carbonyl carbon changing from sp_2 to sp_3 and, consequently, its geometry from planar to tetrahedral. The network of hydrogen bonds involving the Asp38-Ser41-Tyr83-Trp45 residues remained unchanged. No other significant atomistic differences were found when this mechanism was compared to the reaction with the wild type substrate. In this model, this first step was associated with an activation energy of 15.8 $kcal.mol^{-1}$.

4.4.2.2 The second reaction step

To study the second step of this reaction, the distance between the peptidic nitrogen and the Asp226 acidic proton was used as reaction coordinate, similarly to the previously described mechanism. At the transition state (TS2) the proton was between these two residues (distance of 1.30 Å to Leu10 nitrogen and 1.23 Å to the oxygen), being completely transferred only at the products of this step (INT2). The peptidic bond was not completely cleaved at INT2 (bond length of 1.57 Å), the catalytic Asp38 remained deprotonated and the amide nitrogen became positively charged. Therefore, a third reaction step was necessary for the complete hydrolysis of the mutated substrate, as

opposed to the other aspartic proteases and in agreement with the catalytic mechanism for the wild type substrate. This reaction needed an activation energy of 12.0 kcal.mol^{-1}.

4.4.2.3 The third reaction step

The reaction coordinate was taken as the distance from a gem-diol proton to the carboxyl oxygen of Asp38, from where we have obtained the structure of the third transition state and the final products. At the end of the reaction cycle the substrate was hydrolyzed to two independent peptides, in an analogous manner to the wild type substrate. Therefore, the catalytic reaction mechanism of this mutated substrate was similar to the hydrolysis of the natural substrate. Thus, it was concluded that the Leu10Phe mutation did not cause relevant mechanistic changes in the catalytic cycle or structural changes in the enzyme-substrate interaction. The activation energy for the last step was 16.4 kcal.mol^{-1}, which was comparable to the experimental value for this hydrolysis (18.3 kcal.mol^{-1}) [214], and this final step led to a very stable dipeptide (reaction energy of -16.2 kcal.mol^{-1}).
The energy pathway and the representation of the stationary points for this mechanism are given in Supporting Information. The overall conclusion is that the catalytic mechanism of the cleavage of angiotensinogen is the same for the wild type and for the preeclampsia-generating mutant. However, we found differences between the barriers of both mechanisms, based on their PES profiles. Experimental barriers are extremely similar, and the experimental difference is only 0.7 kcal.mol^{-1}. This means that the theoretical method used here identifies the correct chemical pathway, albeit facing challenges in retrieving high-accuracy energies.

4.4.2.4 Human renin and mouse renin

In pre-clinical trials the mouse is the standard animal model to test new drug candidates. However, results on mouse are many times not transferable to humans, due to the different physiologies of the two organisms. Therefore, it becomes very important to compare the catalytic mechanisms of both human and mouse renin, in order to predict how reliable, the mouse models will be. The hydrolysis catalyzed by mouse submandibular renin was previously described in our group [216].
Human renin is 68% homologous in amino acid sequence to mouse renin and the active site of these two enzymes have similar folds. Some differences are observed in only a few residues. In human renin, Ala229 substitutes a threonine residue that is conserved in the most part of the aspartic proteases. This alanine residue lacks the ability to form a

hydrogen bond with an aspartate in the active site. In mouse renin, the same threonine is substituted by Ser226. Another difference is observed in Val36, which is in close proximity to one of the catalytic aspartates in human renin. In the mouse enzyme this residue corresponds to an isoleucine, which is not identical but very similar to the human counterpart [217].

Comparing our present results with our previous studies in mouse, we conclude that both human and mouse renin adopt the same hydrolysis mechanisms, and the same sequence of steps, with similar transition states, to hydrolyze angiotensinogen. We can conclude also that the different residues between these enzymes are not essential to catalysis. In conclusion, the mouse model seems to be reliable to test drug candidates, at least at the level of renin inhibition.

4.5 Conclusions

In this work we have explored, computationally, the mechanism of angiotensinogen hydrolysis by human renin at the atomic level. The results showed that the transformation was accomplished with the help of acid/base catalysts in three elementary steps, contrarily to other proteases of the same family, which perform an identical reaction in just two steps. The first step is characterized by the formation of a gem-diol intermediate upon nucleophilic attack of a water molecule on the substrate carbonyl. The protonation of the peptidic nitrogen of the hydrolyzed bond takes place in the second reaction step. A third and last step is necessary to the formation of the final product, characterized by the transfer of a gem-diol proton back to the Asp 38, simultaneously with the cleavage of the scissile bond between Leu10 and Val11 of angiotensinogen.

The comparison of the results obtained here for the human enzyme with previous works on the mouse enzyme, led us to conclude that these two mechanisms are very similar. As mouse is usually used as model to preliminarily test drug candidates, this comparison gives further confidence about the transferability of the results obtained in this animal model to humans.

The energy pathway obtained for this mechanism showed that the first reaction step is rate limiting, with an activation free energy of 22.0 kcal.mol^{-1} (M06/6-311++G(2d,2p):amber). Similar results were obtained with a set of different density functionals. The experimental value for the hydrolysis of the N-terminal tetradecapeptide of angiotensinogen by renin is 16.5 kcal.mol-1[214]. This is comparable to the value obtained in this work, if we take into account the error associated with the methodology.

The results described in this work also showed that the replacement of a leucine by a phenylalanine, in the position 10 of angiotensinogen, did not change the reaction catalyzed by human renin. Therefore, we can conclude that the association of this mutation with preeclampsia should not be related to changes in the hydrolysis mechanism catalyzed by renin. In fact, the difference between the two enzymes is very small but, together with a change in K_M, the enzyme becomes more efficient, which should play a role in hypertension.

As renin is an essential enzyme for blood pressure control, studies such as the present one is necessary for a better understanding of its role. The complete description of this catalytic mechanism and the atomistic details of its transition state are valuable tools for the rational design of new antihypertensive drugs. The geometry of the transition state for the rate-limiting step can be used as a model for the exploration of new transition-state analogues inhibitors for this enzyme.

4.6 Supporting Information

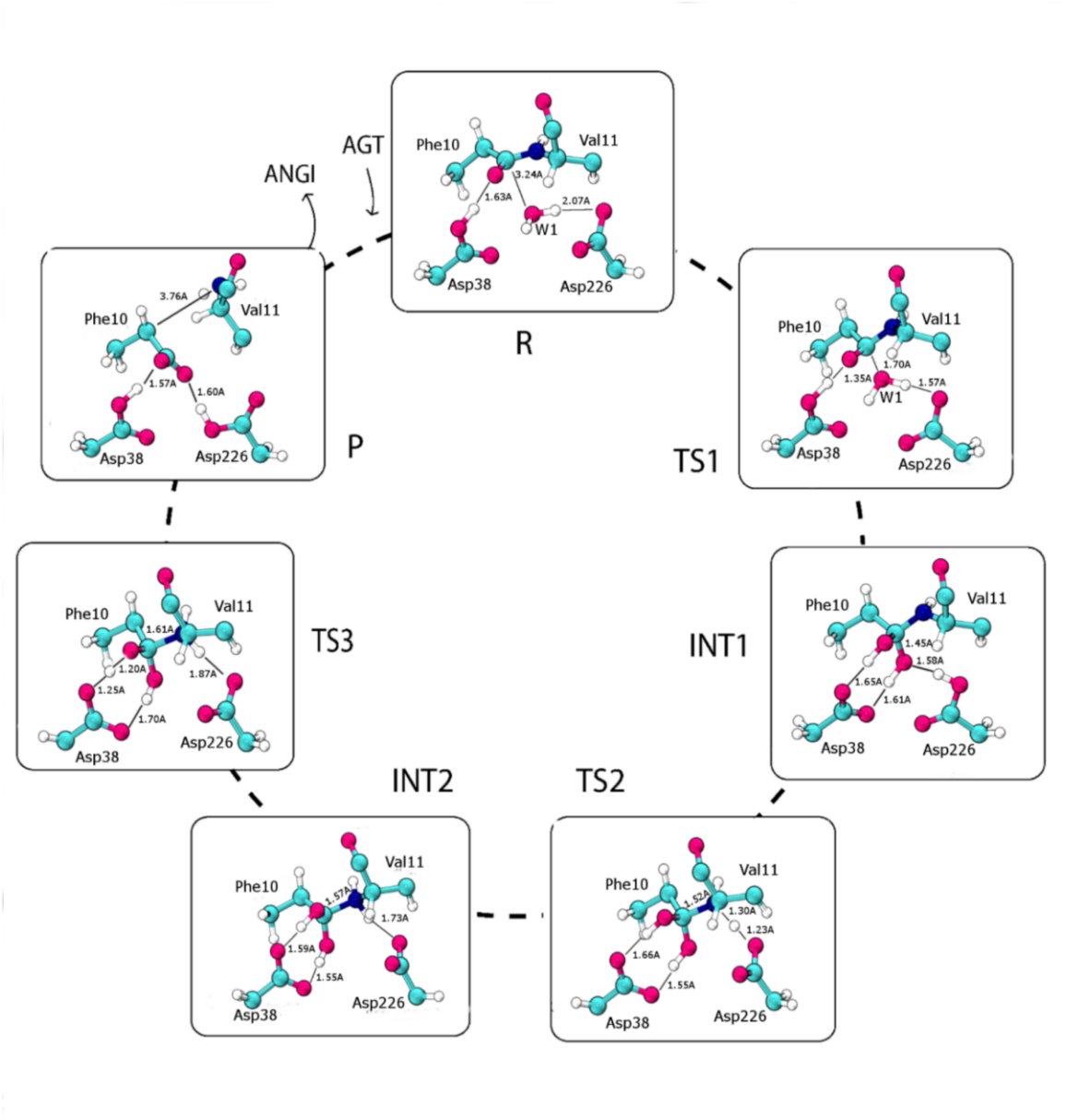


Figure 4.5 (SI) Structures of the reactants, intermediates, transition states and products for the cleavage of the Phe10-Val11 peptide bond of mutated angiotensinogen by human renin. Only the QM-treated atoms are represented in the figure, for the sake of clarity

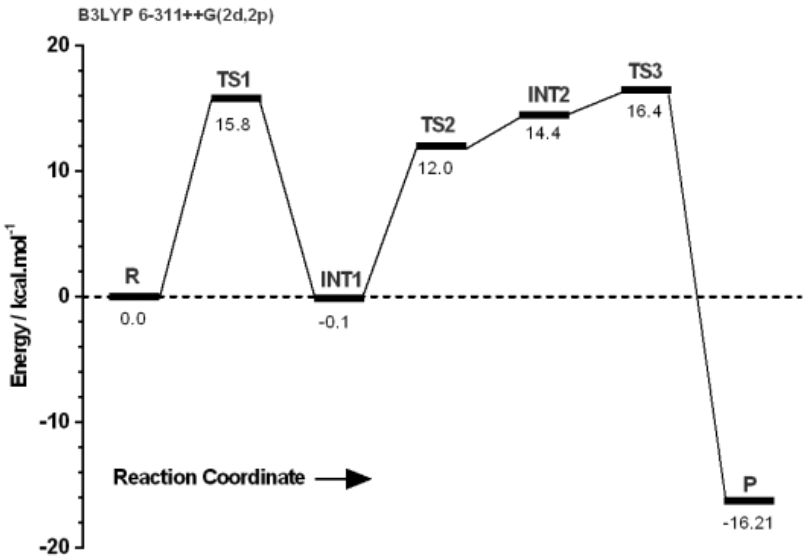


Figure 4.6 (SI) Energetic pathway for the hydrolysis reaction of mutated angiotensinogen catalyzed by human reninat the B3LYP/6-311++G(2d,2p):AMBER level with 139 atoms in the high layer.

Table 4.1 (SI) Activation and reaction energies for angiotensinogen hydrolysis by human renin calculated with density functional and 6-311++G(2d,2p) basis set. These energies included the thermal corrections to Gibbs free energy.

	ΔG /kcal.mol^{-1}						
6-31++G(2d,2p)	**R**	**TS1**	**INT1**	**TS2**	**INT2**	**TS3**	**P**
B3LYP	0.0	25.1	10.5	17.8	19.0	18.4	7.9
M06	0.0	22.0	3.5	13.9	15.5	16.5	5.6
B1B95	0.0	22.1	6.2	13.1	15.0	16.4	7.6
MPWB1K	0.0	21.9	3.7	11.7	13.6	16.2	8.6

CHAPTER 5. The Catalytic Mechanism of the Marine-Derived Macrocyclase PatGmac

Natércia Fernandes Brás[a], Pedro Ferreira[a], Ana R. Calixto[a], Marcel Jaspars[b], Wael Houssen [c,d], James H. Naismith[e], Pedro A. Fernandes[a] and Maria J.Ramos[a]

[a] UCIBIO, REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal; [b] Marine Biodiscovery Centre, Department of Chemistry, University of Aberdeen, Old Aberdeen, AB24 3UE, Scotland, U.K.; [c] Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, Scotland, U.K.; [d] Pharmacognosy Department, Faculty of Pharmacy, Mansoura University, Mansoura 35516, Egypt; [e] Biomedical Sciences Research Complex, University of St Andrews, North Haugh, St Andrews, Fife KY16 9ST, Scotland, U.K.

In this chapter the catalytic mechanism of PatG macrocyclase was described with atomistic detail using QM/MM methodologies. This enzyme is a subtilize-like serine protease that catalyzes the macrocyclization of the patellamides. Considering the ability of this enzyme to catalyze a large range of non-active substrates, this work is not only important to understand how this enzyme works, but it has also a wide range of implications in a large number of areas, such as pharmaceutical and biotechnology industry. Regarding the contributions to the work, Natércia Brás performed the initial calculations. Ana Rita Calixto helps Pedro Ferreira to perform and analyze the final calculations. Pedro Ferreira wrote the first draft manuscript, which was revised through contribution of all authors. This work was published in *Chemistry-a European Journal* and the following content is an integral transcription of the published version.

- Bras, N. F.; Ferreira, P.; **Calixto, A. R.;** Jaspars, M.; Houssen, W.; Naismith, J. H.; Fernandes, P. A and Ramos, M. J., ***The Catalytic Mechanism of the Marine-Derived Macrocyclase PatGmac***. Chemistry-European Journal, 2016, 22 (37), 13089-13097 – Hot Paper **DOI: 10.1002/chem.201601670**

5.1 Abstract

Cyclic peptides are a class of compounds with high therapeutic potential, possessing bioactivities including anti-tumour and anti-viral (including anti-HIV). Despite their desirability, efficient design and production of these compounds has not yet been achieved. The catalytic mechanism of patellamide macrocyclization by PatG macrocyclase domain has been investigated using computational methods. We applied a quantum mechanics/molecular mechanics (QM/MM) methodology, specifically ONIOM(M06/6-311++G(2d,2p):ff94//B3LYP/6-31G(d):ff94). The mechanism proposed here begins with a proton transfer from Ser783 to His 618 and from the latter to Asp548. Nucleophilic attack of Ser783 to the substrate leads on to the formation of an acyl-enzyme covalent complex. The leaving group (AYDG) of the substrate is protonated by the substrate's N-terminus leading to the breakage of the P1-P1' bond. Finally, the substrate's N-terminus attacks the P1 residue, decomposing the acyl-enzyme complex forming the macrocycle. We found that the formation and decomposition of the acyl-enzyme complex have the highest activation free energies (21.1 kcal.mol^{-1} and 19.8 kcal.mol^{-1} respectively), typical of serine proteases. Understanding the mechanism behind the macrocyclization of patellamides will be important to the application of the enzymes in the pharmaceutical and biotechnological industries.

5.2 Introduction

The pharmaceutical industry has become excellent at developing small-molecule drugs (below 600 MW) hitting classical compact binding sites. However, there remain a large number of non-classical targets with extended binding sites which cannot be modulated using small molecules, but instead need biologics (antibodies/native peptides) to obtain a therapeutic effect. Of the top selling drugs on the market 7 out of 10 are biologics aimed at complex diseases such as rheumatoid arthritis and cancer. Macrocycles (500-2000 Da) are a third class of therapeutics which are able to modulate the same complex, extended, targets as biologics, but are easier to administer, hit intracellular targets and may have a lower associated cost of goods [218]. Of the 68 approved macrocycle pharmaceuticals, 27 are cyclic peptides [219], 1 of which is orally available. Cyclic peptides have a number of advantages over linear ones including reduced susceptibility to metabolism, improved membrane permeability and an entropic advantage on binding to a target. The pharmaceutical industry is struggling with two aspects of macrocycles: their design and efficient production. Most processes to generate peptide macrocycles rely on the use of high dilution conditions to prevent oligomerisation, using controlled addition conditions making these approaches non-viable for large-scale synthesis.[220,221]. Alternative synthetic approaches have been developed including on-bead macrocyclization requiring attachment to the solid support via an amino acid side chain and a 3-dimensional orthogonal protecting group strategy [222]. Biological methods of macrocyclization include sortase-mediated ligation, but this results in the incorporation of the pentapeptide LPXTG in the cyclic peptide where X is variable[223].

Efficient formation of cyclic peptides without leaving a residual sequence in the final macrocycle is desirable. Natural cyclic peptides can be formed either via a non-ribosomal peptide synthetase route, or the more recently discovered superfamily of ribosomally produced and post-translationally modified peptides (RiPPs) [222]. RiPPs are formed using a common biosynthesis, in which a precursor peptide, comprised of a leader sequence followed by a core peptide often flanked by signal sequences, is modified via the action of processing enzymes, which install post-translational modifications in the core peptide.

The matured core peptide is then removed from the leader and signal sequences, liberating the mature, active peptide. In many cases, this last step results in the formation of a peptide macrocycle. Three RiPP macrocyclases have been defined; butelase-1 an asparagine/aspartate peptide ligase that is responsible for the formation the plant cyclic peptide cyclotides [224], GmPOPB, a prolyl oligopeptidase involved in α-amanitin biosynthesis [225] and PatG macrocyclase (PatGmac), involved in the biosynthesis of the cyanobactins which are cyanobacterially derived azole/azoline containing cyclic

peptides[226]. The requirements of the PatGmac are a core peptide sequence of 6-11 residues which may include D-amino acids or unnatural amino acid residues, an AYD(GE) signal at the C-terminus which is not incorporated in the final cyclic peptide, and a cyclic residue (proline or thiazoline) at the C-terminal of the core peptide which is included in the macrocycle [227].

Our structural investigations have delineated the mode of action of PatGmac and have led to an understanding of the requirements described above.[226] PatGmac is a subtilisin-like serine protease produced by *Prochloron sp.* which is an obligate symbiont of the seasquirt *Lissoclinum patella* [228]. The more normal extended conformation of the peptide substrate is prevented by bulky enzyme residues (Met660/Arg686/Phe684) protruding into the binding groove as a consequence of a disulphide bridge formation between Cys685 and Cys724. As a result of this only a bent substrate peptide can bind, and this is facilitated by the conformational properties of the proline or thiazoline residue at the P1 position (P3, P2, P1 are the residues before the cleavage site and P1', P2' those after the cleavage site). The unique structural feature of PatGmac is an insertion loop in the normal subtilisin sequence generating a helix-turn-helix motif forming a protective lid over the active site. The AYD signal at the C-terminus of the core peptide binds via the aspartate residue to basic residues in the helix-turn-helix motif (Arg589/Lys594/Lys598). After binding the substrate peptide, a normal serine protease mechanism ensues, but access by water is prevented by the strongly bound AYD signal, thus allowing the amino terminus of the core peptide to loop around and attack the acyl complex, resulting in cleavage of the recognition signal and cyclization of the core peptide.

In this work, we explored the mechanism of the reaction catalyzed by PatGmac using computational methods, in order to describe, with atomistic detail, the most plausible catalytic mechanism of this enzyme. Considering the ability of the enzyme to macrocyclize an extensive range of nonactivated substrates, the data obtained in the present study has wide implications in a number of areas.

Fully understanding the mechanism of PatGmac may lead to engineered analogues with improved properties which accept an even broader substrate range this increasing its utility in a range of pharmaceutical and biotechnology applications.

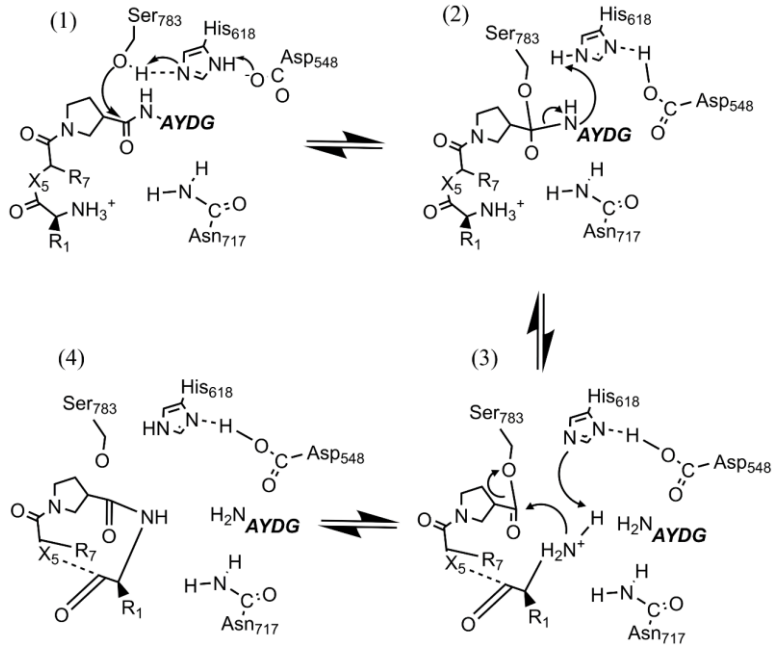


Figure 5.1 Representation of the catalytic mechanism proposed for PatGmac.[226]

5.3 Methods

We started the modelling of the system using an X-ray structure of the subtilisin-like domain of the PatGmac enzyme containing an analogue of the substrate in the active site (PDB: 4AKT, 2.63 Å resolution). This structure had a mutation (His618Ala) on the catalytic triad of the active site (Asp548, His618 and Ser783) which was reverted by superimposition with a structure of the free enzyme (PDB: 4AKS, 2.19 Å resolution) that contained the original residue. The coordinates of Ala618 on the 4AKT structure were then replaced by those of His618 of the free enzyme structure. The hydrogen bonding network between the three residues of the catalytic triad and the proximity between the deprotonated nitrogen of the imidazole ring of His618 and the hydroxyl group of Ser783, observed on the free enzyme structure, is mandatory to generate a productive conformation and to initiate the catalytic mechanism. Thus, we naturally assumed that the rotameric state and position of His618 is similar in the free enzyme and in the enzyme complexed with the substrate. Additionally, in the X-ray structure, a loop composed by residues 651-657 was missing and was modelled using the program MODELLER[229]. This loop is located away from the active site (> 10 Å), thus, the modelling performed should not have a significant effect on the study of the catalytic mechanism. One natural substrate has the sequence Ile-MeOxH-Ala-ThH-Ile-OxH-Phe-ThH-Ala-Tyr-Asp-Gly, the (Ala-Tyr-Asp-Gly) are the recognition residues cleaved (at the ThH-Ala bond) during the reaction (ThH = thiazoline, OxH = oxazoline, MeOxH = methyl oxazoline). The mimic peptide on

the 4AKT structure had a different sequence (Val-Pro-Ala-Pro-Ile-Pro-Phe-Pro-Ala-Tyr-Asp-Gly) in which the azoline heterocycles had been replaced by Pro. Hence, the heterocycles that on the X-Ray structure were mimicked by prolines, were corrected and missing parts were modelled in GaussView [230]. We have preserved the original coordinates for most of the residues of the substrate. The original precursor peptide and the modelled peptide are represented in Supporting Information. We used the X-leap [110] software to protonate the complex, assuming that all residues were in their physiological protonation states, and 23 Na+ counter-ions were added to neutralize the charge of the system. Additionally, we surrounded the system with 15830 water molecules using a truncated rectangular box of TIP3P water molecules with a minimum distance of 12 Å between any atom of the protein and the faces of the box.

We performed a two-step minimization using the Amber 12 [110] simulation package (parm 99SB force field) in order to relax the system by removing eventual tensions and clashes. First, the water molecules and counter-ion were minimized with the remainder of the system fixed (steepest descent algorithm for the first 500 cycles, and conjugate gradient algorithm for the last 1500 steps); and, second, the position of all atoms (steepest descent algorithm for the first 5000 cycles, and conjugate gradient algorithm for the last 10000 steps) of the model was minimized.

We ran molecular dynamics (MD) simulations, starting from the structure obtained after the minimization procedure, to see if the modelled structure was stable and preserved. First we warmed the system from 0 to 300 K in a 200 ps long simulation maintaining a constant volume and using periodic boundary conditions.

Then, a 15 ns production run was conducted using periodic boundary conditions with the isobaric-isothermal ensemble defining a pressure of 1 atm and a temperature of 300 K using the Langevin thermostat and the Berendsen barostat for that purpose. The cutoff for the Lennard-Jones interactions was set to 10 Å. The Coulomb interactions were treated using the Particle Mesh-Ewald (PME) method, with a cutoff of 10 Å for the real part of the sum. The time step of the simulation was 2 fs. The potential energy of the system was treated with the LeapFrog integration algorithm.

A structure of PatGmac complexed with substrate was taken from the MD simulation and was used as the starting point for the study of its reaction mechanism. We have chosen a structure of the system (after the equilibration) whose conformation was productive for the beginning of the catalytic cycle, which means that the distance between Ser783 and the substrate, and between the substrate's N-terminus and P1 had to be small and appropriate for chemical reaction. The analysis of the trajectories has shown that, after the equilibration, these criteria are met in most of the structures, hence, the entropic cost for reaching a productive conformation should not be significant. To investigate the

potential energy surface (PES) along the mechanism of the macrocyclization reaction we performed QM/MM calculations, widely used in enzymatic studies [199,231-233] applying an ONIOM [132] scheme as implemented in the Gaussian 09 software package [135]. The system, containing a total of 5200 atoms, was divided into a "QM layer", containing 91 atoms (**Figure 5.2**)**,** and an "MM layer" which were treated at DFT and classical MM levels respectively. The high layer includes the catalytic triad (Ser783, His618 (side chain until Cα) and Asp548 (side chain until Cβ)), Asn717 (side chain until Cγ) and P1 (Cys), P1' (Ala), P2' (Tyr) residues and the terminal Ile of the substrate. A list of the atoms in the QM layer is given in table 5.1 (SI). For the QM layer we employed the B3LYP functional using 6-31G(d) basis set for geometry optimizations, which was shown to provide accurate results in previous studies [71,216,234,235] while in the low layer we used the AMBER parm 99SB force field. We used hydrogen atoms as link atoms where covalent bonds were in between the two layers. The interaction between the two layers was treated using an electrostatic embedding scheme. The study of each mechanistic step began by conducting linear scans along the reaction coordinates. These corresponded to specific interatomic distances that connected the reactants to the products of each hypothesized reaction step. The precise reaction coordinates that were assumed are described in the main text, when each reaction step is discussed.

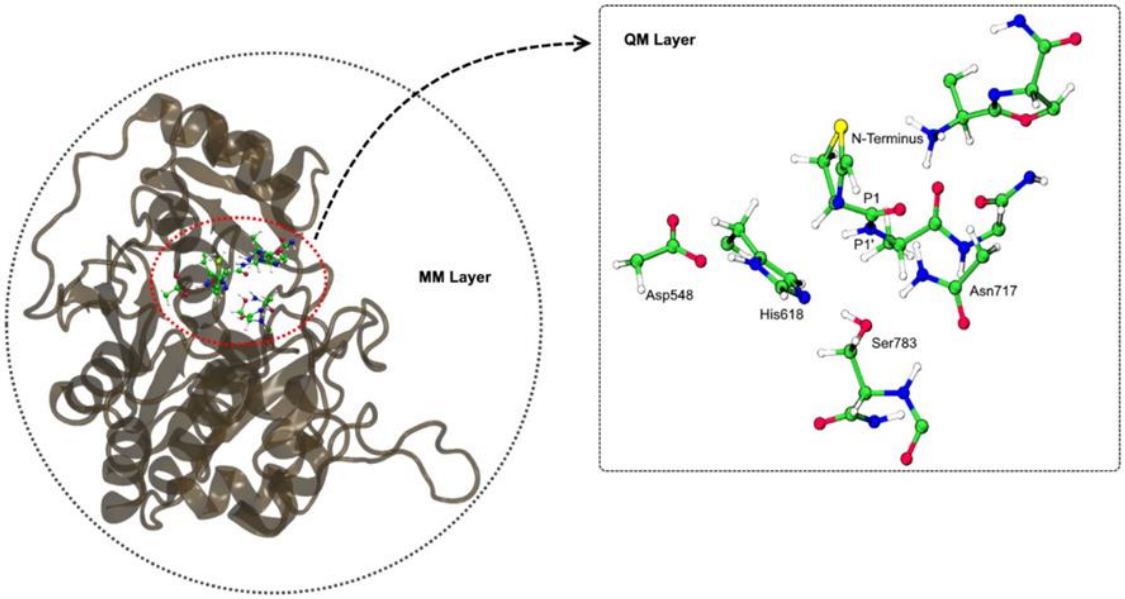


Figure 5.2 Model used in the QM/MM calculations. The high layer is composed by 91 atoms, which includes the residues of the catalytic triad (Ser783, His618 (side chain until Cα) and Asp548 (side chain until Cβ)), Asn717 (side chain until Cγ) and by P1(Cys), P1'(Ala), P2'(Tyr) residues and the terminal Ile of the substrate.

The structures of the reactants, intermediates, transition states (TS) and products were then fully optimized, starting with the guesses taken from the linear transit scans, for the rate-limiting steps. In the case of chemical steps with very shallow barriers we have used

the highest energy structure of the linear transit scan as very good approximations of the transition state. This procedure was motivated by the great complexity of performing free transition state geometry optimization in this very large and heterogeneous system. The differences in free energy that we got when we compared the two procedures (i.e. taking the structure from the adiabatic mapping and making a free geometry optimization, both alternatives calculated for the rate-limiting steps) were 0.4 and 2.6 kcal.mol^{-1}. We performed vibrational frequency calculations for every stationary point, with those of the reactant, intermediates and product having no imaginary frequencies and those of the TS having just one, which in all cases was clearly related to the reaction coordinate. Even though the identity of the minima connected to each transition state was clear from the liner transit scans, and was further supported by the observation of the relevant normal mode, we ran further IRC calculations (albeit not to the full extent of the explored PESs), starting from the obtained TSs, to confirm that the minima that were connected to the TS were the ones that we were expecting. The final energies were obtained conducting single-point (SP) energy calculations on the optimized geometries using different density functionals (M06, B1B95 and mPWB1K in addition to B3LYP), known to have a good performance for thermodynamics and kinetics, and a larger basis set, 6-311++G(2d,2p), in the QM layer to improve the accuracy of the results. Table 5.2 (SI) shows the activation and reaction energies obtained with the different functionals. They provide a PES that translates into the same mechanism and are qualitatively equivalent. The values obtained with B3LYP seem a bit elevated comparatively to those obtained with the other functionals. The energy barriers for each step are similar in all cases. However, with B3LYP and M06 the energy barrier to TS3 is slightly higher than the barrier to TS5 whereas with mPWB1K and B1B95 the opposite is observed. In the discussion we considered the results of the M06 functional because it has been shown to be the most appropriate for the description of the thermodynamics and kinetics of the chemistry of main-group elements [121,236,237]. We have calculated also the zero-point energy, the entropy and the thermal corrections to obtain Gibbs free energies at 298.15 K, which is a comparable temperature to that of the water where the Indo-Pacific seasquirt Lissoclinum patella naturally occurs.[238] To calculate the entropy and free energy we have employed the particle in a box/rigid rotor/harmonic oscillator formalism. This is a physically clear and rigorous formalism to calculate the entropy and free energy of a system within a single-conformation model.

GD3 dispersion [123] was included in the single point calculations as implemented in Gaussian 09 D.01.

5.4 Results and Discussion

In previous work, some of us proposed a catalytic mechanism for PatGmac [226] **(Figure 5.1**) based on the structure of the enzyme macrocyclase domain and biochemical characteristics. It was suggested that the first step of the reaction is the nucleophilic attack of Ser783, aided by His618 which acts as a base, to the P1' carbonyl group of the substrate leading to the formation of an enzyme-substrate tetrahedral intermediate. After this first attack, the AYDG peptide is cleaved by protonation of the leaving group's N-terminal by His618. Lastly, the N-terminus of the substrate attacks the carbonyl group of the cysteine (P1) forming the macrocycle. In order to test the proposed mechanism and get a complete and detailed description of this catalytic mechanism, we have performed QM/MM calculations.

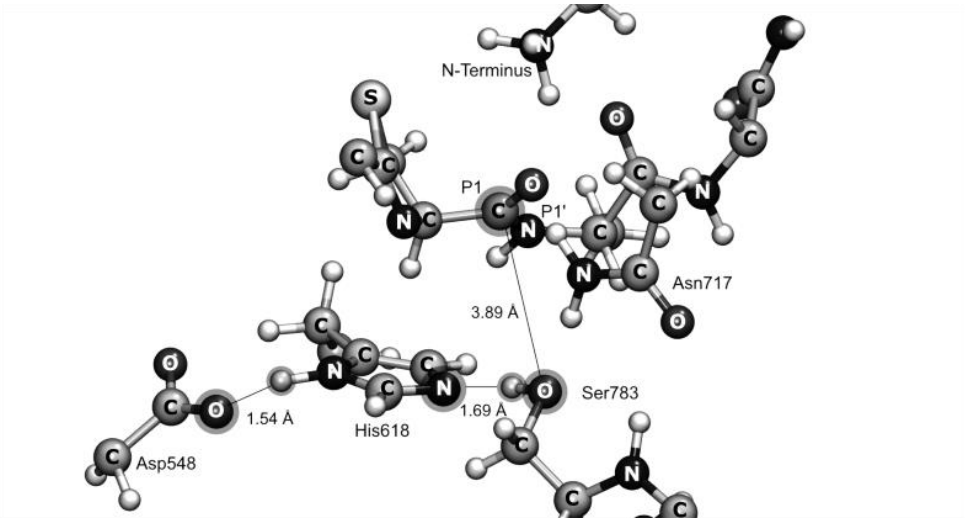


Figure 5.3 Structure of the reactants highlighting the most relevant distances for the first mechanistic step.

5.4.1 The first reaction step

In the reactants optimized structure, Ser783 is at 3.89 Å from the carbon atom of P1 and correctly oriented to begin the attack. Asp548 is hydrogen bonded to His618 (1.69 Å) and the latter is hydrogen bonded to Ser783 (1.54 Å) (**Figure 5.3**). The catalytic reaction begins with the typical proton transfer on the serine protease's catalytic triad, from Ser783 to His618. We defined the distance between $N_{\delta 1}$ of His618 and the H_{γ} of Ser783 as the putative reaction coordinate. As the proton of Ser783 was being transferred to His618, the proton of the His618 imidazole ring spontaneously moved to Asp548 O^{-} atom. We noticed that the structures of the reactants (Asp-COO^{-}/H-His/OH-Ser) and of the generated

intermediate (Asp-COOH/His-H/-O-Ser) are very close in energy (2.4 kcal.mol^{-1}). The free energy barrier was calculated in 2.5 kcal.mol^{-1}. At the transition state structure, both the proton transferred from Ser783 to His618 and the one that was consequently transferred from His618 to Asp548, assume an intermediate position between the respective residues **(Figure 5.4).**

The vibrational frequency analysis revealed an imaginary frequency corresponding to the reaction coordinate (an antisymmetric stretch involving both proton transfers) confirming the nature of the transition state (996.52 *i*.cm^{-1}). However, another small negative frequency (31.30 *i*.cm^{-1}) was found, as a consequence of the more relaxed geometry optimization.

Thermal corrections and zero point energies were included in the free energy calculation, despite the existence of a very small second imaginary constant. These corrections have a small contribution to the corresponding free energy values in agreement to what has been reported in the literature regarding proton transfers between catalytic residues (ca. 1.2 kcal/mol) [239]. As the barrier was extremely shallow and not rate-limiting it was not optimized further.

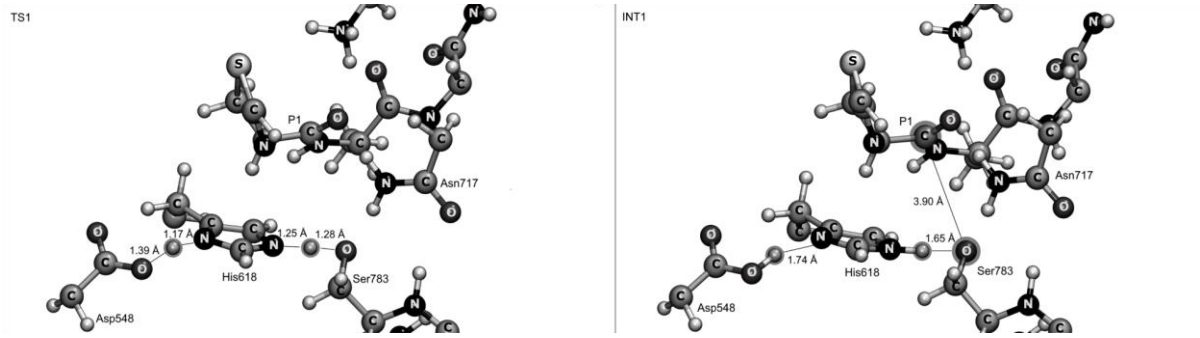


Figure 5.4 Representation of the structures of the first transition state (TS1) and the intermediate (INT1) on the first step

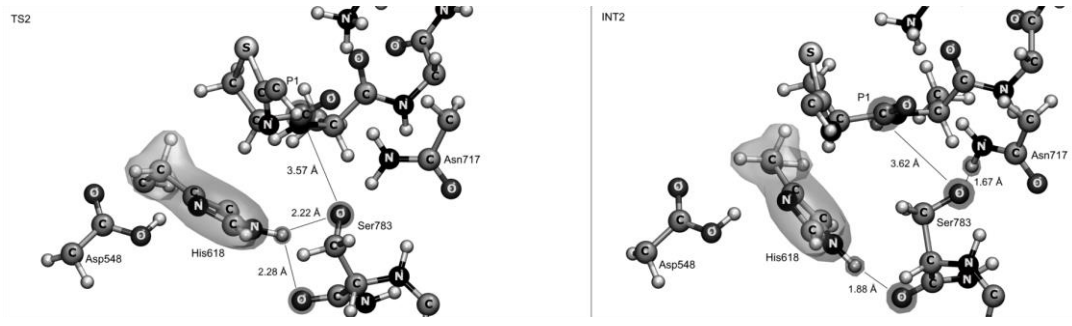


Figure 5.5 Representation of the structures of the transition state (TS2) and the intermediate (INT2) of the second step of the macrocyclization reaction

The small free energy of reaction suggests that both structures may exist simultaneously being the latter the most favorable to initiate the catalytic reaction. The possibility of the

first step to begin with a direct nucleophilic attack of Ser783 to the substrate, with the proton transfer occurring during the attack was also considered. Defining the distance between Ser783 and P1 as reaction coordinate, it was found that the energy barrier associated with this alternative is very high (>30 kcal.mol^{-1}). Our proposal for the first step has a much smaller energy barrier intermediate with enhanced nucleophilicity for attack at the substrate amide.

5.4.2 The second reaction step

The deprotonated Ser783 attacks the P1 carbonyl group forming the enzyme linked tetrahedral intermediate. A study of this step was made with a linear transit scan using the optimized structure after the proton transfer described above as reactants, and defining the distance between the O_γ of Ser783 and the carbonyl carbon of P1, which is initially 3.90 Å (**Figure 5.4**), as the reaction coordinate to be scanned.

A conformational change in His618 took place before the formation of the tetrahedral intermediate, with the approximation of Ser783 to the P1 carbonyl group, His618 changed its rotamer by rotating around the C_β-C_γ bond, breaking the hydrogen bond with Ser783 hydroxyl and forming another with the Ser783 carbonyl, while keeping the hydrogen bond with Asp548. Additionally, a hydrogen bond between the attacking oxygen of Ser783 and Asn717 amine was formed (**Figure 5.5**). This conformational change led to a stable intermediate (INT2) before Ser783 has completed the nucleophilic attack on the substrate. A free energy barrier of 4.8 kcal.mol^{-1} in relation to INT1 was found, characterized by one imaginary frequency (31.83 *i*.cm^{-1}) clearly corresponding to the reaction coordinate. We will refer to this point as TS2 (**Figure 5.5**). In that structure His618 is placed between Ser783 side chain and carbonyl oxygen atoms at 2.22 Å and 2.28 Å respectively. Ser783 which initially was at 3.90 Å from the P1 residue stays at 3.57 Å on TS2 and at 3.61 Å on INT2.

5.4.3 The third reaction step

Starting from the INT2 optimized structure, the attack of Ser783 to the substrate was conducted until the formation of the tetrahedral intermediate, using the distance between the Ser783 deprotonated oxygen and the P1 carbonyl carbon (3.62 Å) as reaction coordinate (**Figure 5.5**). This attack had a free energy barrier of 21.1 kcal.mol^{-1} in relation to INT2. This rate-limiting transition state (TS3) was further freely optimized, having one imaginary frequency (744.16 *i*.cm^{-1}) that corresponds to the coordinated stretching of the

Ser783 oxygen-P1 carbonyl carbon and of the N-H bond of the terminal amino group of the substrate.

This proton, together with the Asn717 amine, constitutes the oxyanion hole which stabilizes the oxyanion of the acyl-enzyme complex. The interactions between them may consist of low-barrier hydrogen bonds (LBHB) considering, particularly, their short length (< 2.0 Å). The existence of LBHB between the oxyanion and the oxyanion hole in serine proteases has already been proposed [240] although others have considered them to be simple electrostatic interactions[241]. In this case we can clearly see the transition from an electrostatic hydrogen bond in the reactants to a hydrogen bond where the proton is shared between both acceptors, commonly known as a single-well hydrogen bond, which corresponds to an extreme case of a low barrier hydrogen bond where the barrier vanishes. However, we note that this structure, achieved at the transition state, is not a stable, long-lived interaction. Whether or not this kind of interactions are catalytic or anticatalytic is a matter of debate [20].

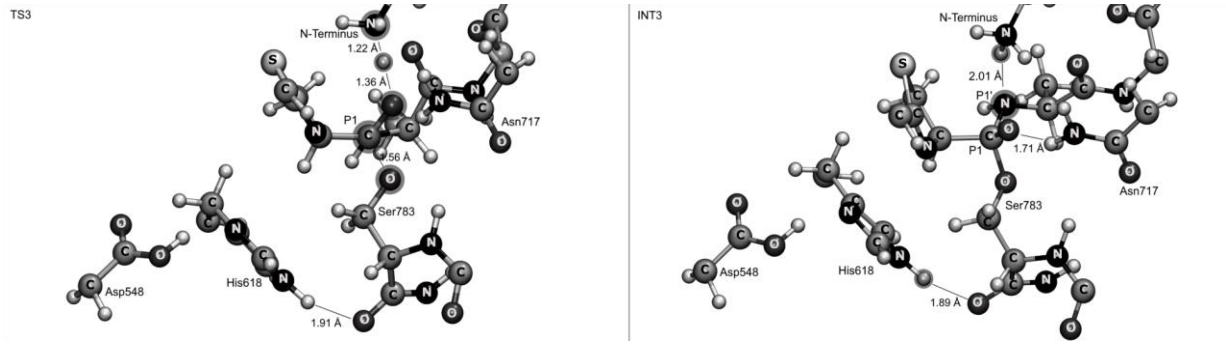


Figure 5.6. Representation of the structures of the transition state (TS3) and the intermediate (INT3) of the third step of the reaction.

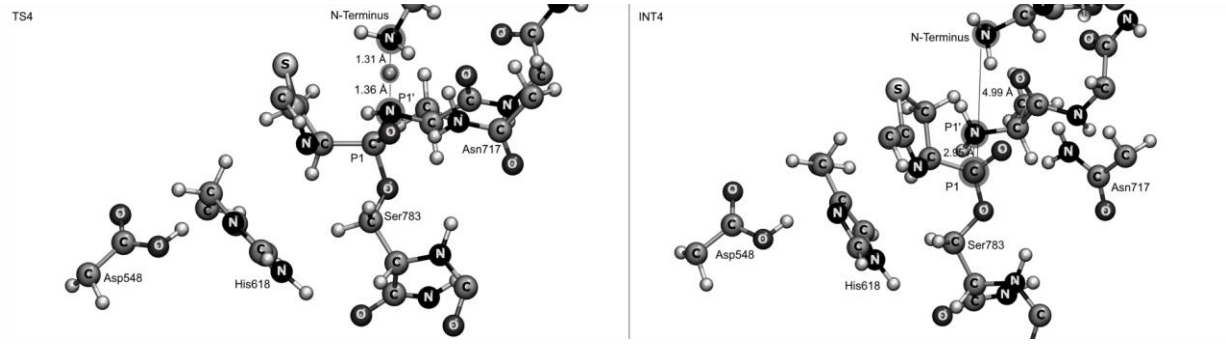


Figure 5.7 Representation of the structures of the transition state (TS4) and the intermediate (INT4) of the fourth step of the reaction.

5.4.4 The fourth reaction step

The conformational rearrangement that His618 makes it unlikely that it protonates the leaving group in the third step of the catalytic reaction, as initially proposed and which is typical for other serine proteases [241]. After the rearrangement His618 stays at 5.97 Å from P1' and not adequately oriented (**Figure 5.6**), the substrate's N-terminal (charged) amine group would be the most suitable candidate to protonate P1' amine.

Therefore, we conducted a linear transit scan using the distance between the $NH3^+$ terminal proton and the P1' amine group, which is 2.01 Å in the INT3 geometry, as reaction coordinate (**Figure 5.6**). A transition state (TS4) was then freely optimized and an imaginary frequency (1184.98 *i*.cm^{-1}) was found which corresponds to an asymmetric stretch between the transferred proton and the two nitrogen atoms (**Figure 5.7**). We found an energy barrier of 3.1 $kcal.mol^{-1}$ relatively to INT3. With the transfer of the proton, a very stable intermediate (INT4) is obtained (ΔGstep4 = - 23.8 $kcal.mol^{-1}$), the bond P1-P1' is cleaved but the AYDG tetrapeptide is retained on the active site of the enzyme, at 2.95 Å from P1, on the INT4 optimized structure (**Figure 5.7**)

5.4.5 The fifth reaction step – Macrocyclization of the substrate

To complete the catalytic mechanism of PatGmac, the substrate's N-terminus attacks the carbon atom of P1 carbonyl group closing the macrocycle. We have used the distance between the N-terminus and the P1 carbonyl carbon (4.99 Å) as reaction coordinate. We found that during this step the N-terminus donates a proton to Ser783, consequently decomposing the enzyme-substrate tetrahedral structure since the bond between P1 and Ser783 is cleaved. The macrocycle is formed and the AYDG peptide which was at 2.95 Å from P1 at the beginning of this step is further displaced. This step had a free energy barrier of 19.8 $kcal.mol^{-1}$, in relation to INT4, and originates very stable products (ΔGstep5 = - 42.6 $kcal.mol^{-1}$) (**Figure 5.7**), being the displacement of the AYDG peptide an important factor contributing to the great stability of the products. The freely optimized TS5 geometry shows one imaginary frequency (136.59 *i*.cm^{-1}), corresponding to the reaction coordinate. The present description of this step differs from the first proposed mechanism [226] ,in that, the substrate's N-terminus donates a proton to Ser783 rather than His618.

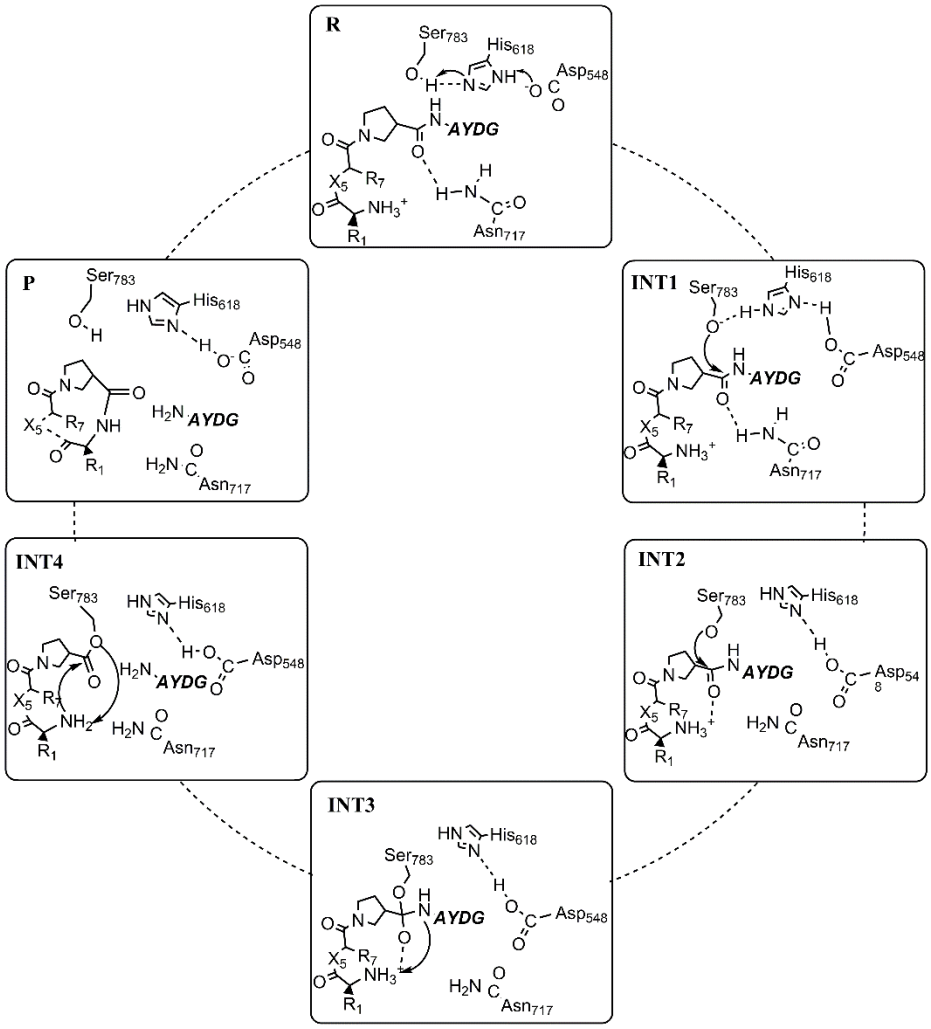


Figure 5.8 General scheme for the catalytic mechanism of PatGmac predicted by earlier experiments and by present QM/MM calculations.

5.4.5 Energetic profile of the PatGmac catalytic mechanism

Figure 5.8 shows the energetic pathway of the macrocyclization reaction catalyzed by PatGmac at the ONIOM(M06/6-311++G(2d,2p):Amber//B3LYP/6-31G(d):Amber) level. According to these results, the TS2 has the highest free energy. However, that is not the rate-limiting step since both TS3 and TS5 have higher energy barriers relatively to the preceding intermediate which are very stable. These TS represent the formation of the enzyme-substrate complex (TS3) and the formation of the macrocycle and accompanying decomposition of the acyl-enzyme complex (TS5). The TS3 has an energy barrier slightly higher than TS5 (21.1 kcal.mol^{-1} and 19.8 kcal.mol^{-1} respectively), but given their proximity, both will be relevant for the observed rate of this reaction (note that the difference between the two is very narrow and probably close to the accuracy of the methodology for relative energies between similar molecular structures). Both the

formation and decomposition of the tetrahedral complex have been identified as slow steps previously [241,242] consistent with our results here and with our previous mass spectrometric observation of the acyl PatGmac intermediate. The PES of the reaction shows that it is strongly exothermic with the products 76.0 kcal.mol^{-1} more stable than the reactants. All mechanistic steps are exothermic with step 3 (formation of the acyl-enzyme intermediate) being the sole exception.

The reaction rates reported for PatGmac are approximately 1 per day.[227,243.] Thus, the Gibbs free energy barrier for the macrocyclization reaction may be estimated from the transition state theory, resulting in an observed experimental free energy of ≈ 24 kcal.mol^{-1}, which is a comparable value to the obtained in the present work (21.1 kcal.mol^{-1}).

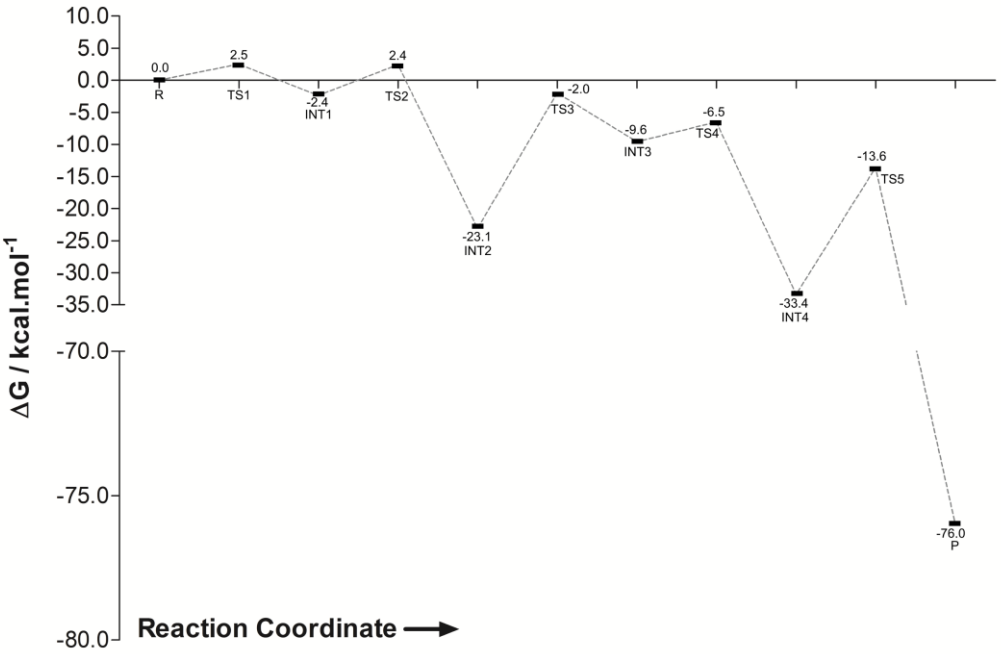


Figure 5.9 Potential energy surface (PES) for the macrocyclization reaction catalysed by PatGmac. The energies were obtained at ONIOM(M06/6-311++G(2d,2p):Amber//B3LYP/6-31G(d):Amber) level.

5.5 Conclusions

We have explored by computational approaches the macrocyclization reaction catalyzed by PatGmac. Our results showed that the mechanism followed by this enzyme is different to those typical of serine proteases albeit with some similarities (Scheme 1). The typical proton transfer on the Ser-His-Asp catalytic triad occurs, as we describe in the first reaction step, as well as the formation of the acyl-enzyme intermediate (second mechanistic step). We found, however, that the protonation of the AYDG leaving group of

the substrate is most probably made by the substrate's $NH3^+$ terminus and not by His618 (**Figure 5.1**), due to steric impediments (**Figure 5.2**). This enzyme differs from typical serine proteases where deacylation of the enzyme-substrate complex occurs by an attack of a water molecule regenerating the enzyme. In PatGmac, the active site is shielded from water [226] and because of that, the deacylation is achieved by the attack of the substrate's N-terminus which also protonates the acyl complex. The formation of the peptide bond is common to macrocyclization of other peptide substrates, namely other cyanobactins [226,244].

This new mechanism differs from that proposed previously **Figure 5.1** the central difference is His618 undergoes a conformational rearrangement and does not protonate the leaving group. Rather it is the incoming substrate amino terminus that protonates the leaving group.

This study contributes to further understanding the mechanism of macrocyclization of PatG substrate. As cyclic peptides have been seen as having great interest for industry, particularly pharmaceutical, more knowledge about their natural synthesis also contributes to improve the efficiency of large scale production of such compounds. The findings in this work suggest that adapting the enzyme process to utilize different substrates would need careful consideration of the pKa of the incoming nucleophile. Regarding the active site residues, the Asp His Ser triad is mandatory for the function of the enzyme and have to be maintained in any engineered analogue while Asn717 stabilizes the tetrahedral-intermediate. Viable options to increase the reaction rate have to focus on the rate-limiting step. For so, the most promising way to achieve catalysis would be to optimize the oxyanion hole. Possibly the presence of a positively charged residue at position 717 (i.e. Lys) would further favor the nucleophilic attack of Ser783 to the substrate which we found that it is a rate limiting step of the reaction. However, the size of Lys and the proximity with the positive N-terminus are obstacles for the correct placement and protonation of the lysine.

5.6 Supporting Information

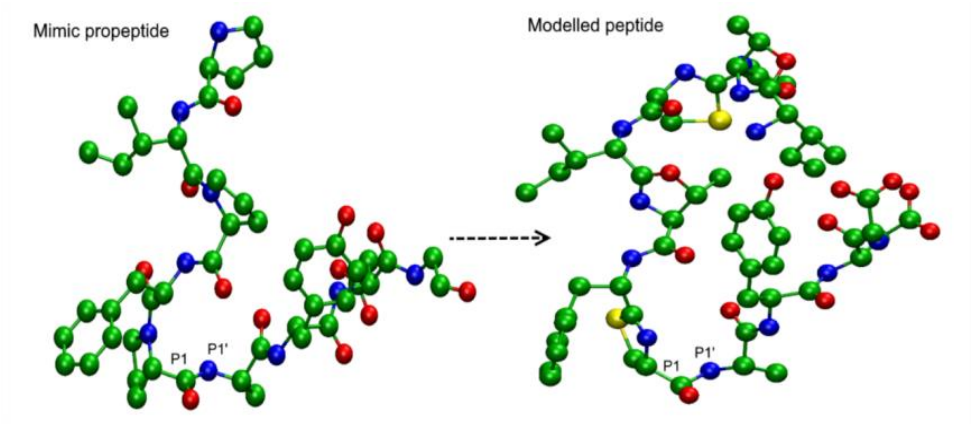


Figure 5.10 (SI) 3D representation of the mimic precursor peptide and of the modelled peptide.

Table 5.1(SI) List of atoms included on the QM layer for the QM/MM calculations.

Residue	Atoms in High layer
P1 (Cys)	CO-C_αH-N-CβH_2-S
P2 (Phe)	C-C_αH
P7 (Thr)	N-C_αH-CO-C_βH
P8 (Ile)	NH_3^+ -C_αH-CO-C_βH
P1' (Ala)	NH-C_αH-CO-$C_\beta H_3$
P2' (Tyr)	NH-C_αH-CO
Asp548	$C_\beta H_2$-C_γOO
His618	C_αH-$C_\beta H_2$-C_γ-Nδ_1H+-Cε_1H-Nδ_2-Cε_2H
Asn717	$C_\beta H_2$-$C_\gamma O_{\delta 1}$-N$\delta_2 H_2$
Thr782	CO
Ser783	$C_\beta H_2$-$C_\gamma O_{\delta 1}$-N$\delta_2 H_2$
Met784	NH

Table 5.2 (SI) Activation and reaction energies obtained for every intermediate, transition state and products of the macrocyclization reaction, obtained with four density functionals and 6-311++G (2d,2p) basis set. Thermal corrections and zero point energies were calculated for every stationary point or transition state. * The unique exceptions are the structures of TS1 and TS2 that weren't successfully optimized to a transition state but frequencies calculations on the highest energy structure of the respective PES scans were carried out. GD3 dispersion was included on the calculations as implemented in Gaussian 09 D.01

	ΔG /kcal.mol^{-1}									
6-31++G(2d,2p)	**TS1***	**INT1**	**TS2***	**INT2**	**TS3**	**INT3**	**TS4**	**INT4**	**TS5**	**P**
B3LYP	1.0	-1.4	4.9	-20.3	3.3	-3.1	-1	-27.7	-10.2	-72.8
B1B95	1.8	-1.8	4.1	-20.4	0.7	-7.5	-4.2	-30.3	-9.4	-72.5
M06	2.5	-2.4	2.4	-23.1	-2.0	-9.6	-6.5	-33.4	-13.6	-76.0
MPWB1K	3.1	-1.3	4.7	-19.9	0.2	-7.8	-3.2	-32.1	-6.3	-73.8

CHAPTER 6. Influence of frozen residues on the exploration of the PES of enzyme reaction mechanisms.

Ana Rita Calixto, Pedro Alexandrino Fernandes and Maria João Ramos

UCIBIO, REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal,

To study the origin of the catalytic power of enzymes, one of the strategies was to explore the catalytic mechanism of different enzymes, starting by different initial structures, evaluating how the potential energy surface fluctuates with time and how protein motions influences the catalytic mechanism of enzymes. These works were described in Chapter 7 and 8, however, before that, we also explore the best way to represent the systems that we choose to study. In the following work, we evaluate the influence of fix some residues in a given radius surrounding the active site. This approximation is usually applied in the exploration of an enzyme reaction mechanism by computational methods. Before to go forward with the exploration of enzyme mechanism with large sampling of initial structures, we decide to explore this methodologic issue, to prepare more consciously the computational models to be used in the succeeding works.

Regarding the contributions to this work, Ana Rita Calixto performed all the calculations and wrote the first draft manuscript, which was revised through contribution of all authors. This work was published in the *Journal of Chemical Theory and Computation*, and the following content is almost an integral transcription of the published version.

Calixto, A. R.; Ramos, M. J. and Fernandes, P. A., ***Influence of Frozen Residues on the Exploration of the PES of Enzyme Reaction Mechanisms***. Journal of Chemical Theory and Computation, 2017, 13 (11), 5486-5495 **DOI: 10.1021/acs.jctc.7b00768**

6.1 Abstract

In this work, we studied one of the very widely used approximations in the prediction of an enzyme reaction mechanism with computational methods, i.e. fixing residues outside a given radius surrounding the active site. This avoids the unfolding of truncated models during MD calculations, avoids the expansion of the active site in cluster model calculations (albeit here only specific atoms are frozen), and prevents drifting between local minima when adiabatic mapping with large QM/MM models is used.

To test this, we have used the first step of the reaction catalyzed by HIV-1 protease, as the detrimental effects of this approximation are expected to be large here. We calculated the PES with shells of frozen residues of different radii. Models with free regions under a 6.00 Å radius showed signs of being overconstrained. The QM/MM energy barrier for the remaining models was only slightly sensitive to this approximation (average of 0.8 kcal.mol-1, maximum of 1.6 kcal.mol-1). The influence over the energy of reaction was almost negligible. This widely used approximation seems safe and robust. The resulting error is on average below 1.6 kcal.mol^{-1}, which is small when compared with others deriving from e.g. the choice of the density functional or semiempirical MO/SCC-DFTB method, the basis set used, or even the lack of sampling or incomplete sampling.

6.2 Introduction

Enzymes are essential molecules due to their unique capability to catalyze biochemical reactions by many orders of magnitude with high specificity and under physiological conditions. Understanding how enzymes are capable to do that is one of the most important questions in biochemistry. The answer to this question is very important not only as a fundamental demand to understand chemical reactions in life but also to contribute to a range of biotechnological applications [24,25,39,164,245,246]. The large size and complexity of enzymes make them difficult to study. Experimental methods, such as spectroscopic techniques, mutagenesis or kinetic studies, have provided many structural, conformational and kinetic information of enzymatic processes. In turn, computational methods are able to offer the atomic resolution needed to describe these catalytic reactions that is mostly unachievable by experiments. However they require approximations to get a good balance between accuracy, time and the computational power needed to describe the chemical rearrangements on enzymes, in order to calculate accurately their (free) energy, and to explore their many plausible catalytic mechanisms [88]. Combined quantum mechanics / molecular mechanics (QM/MM) approaches have become one of the methods of choice to model these enzymatic reactions. These methods, which were first developed by Warshel and Levitt in 1976 [126] have been extremely used to describe a large number of enzymatic reactions [89,131,133,143,154,247-253]. Other choices such as the cluster model [254,255], the empirical valence bond method [150,152,256], Carr-Parrinelo MD [64,149] or metadynamics [257] have found also wide applicability in this field. The object of this study, *i.e.* the consequences of freezing shells of outer residues, is potentially applicable to all of them.

6.2.1 QM/MM methods to model enzyme catalyzed reaction mechanisms

The fundamental principle of QM/MM methods is simple [126,258]. The system is divided in two or more regions: a smaller region that is treated with quantum mechanics (QM), and a large region that is described by empirical molecular mechanics (MM). The QM region should include at least the chemical groups that are directly involved in the breaking and making of chemical bonds, or establish first shell interactions with them, and the MM region should include the remaining residues, which are not directly involved in the chemical reaction.

When modelling an enzyme reaction mechanism using single X-ray conformation protocols, these methodologies find first an approximated minimum energy path (an

enthalpic profile) from the reactants to the final products, for each reaction step, with subsequent unconstrained optimization of the transition states and minima (apart from the outer shell of frozen residues). Rotational / vibrational entropy is later added to get the free energies associated with the studied reactions. If conformational sampling is considered (often with lower level QM Hamiltonians) then a potential of mean force is calculated along a pre-defined reaction coordinate of lower dimensionality. Decades of calculations have shown that none of the protocols performs better than the other. Instead, they complement each other; they focus on overcoming different limitations of the calculations and will converge into a common methodological framework in the future, when high levels of theory can be combined with extensive sampling of the configurational space.

In this work, we will focus on adiabatic mapping techniques even though the conclusions are mostly transferable to the protocols based in MD potentials of mean force.

Apart from the important question of sampling, other approximations are commonly used, which have not been addressed in this study. They consist in the way in which the boundary between regions is treated, the approach used to calculate the interaction between the QM and MM regions (mechanical, electrostatic or polarized embedding), the accuracy of the Hamiltonians used to describe each region, the size of the QM region, or the specific definition and the dimensionality of the postulated reaction coordinate, among others. The quality of the enzymatic model (structural data) and the assignment of protonation states also influence the results.

During the last years a significant number of enzymatic catalytic mechanisms has been described and published, not only by our group [212,234,259-261], but also by many other groups [262-268], employing QM/MM methods, in particular the ONIOM method [132,269]. Many times, due to the complexity of the studied system or to accelerate the calculations, a truncation and/or a group of constraints are applied to the residues in the MM region. In most of them, different shells with different number of residues are frozen during the calculations, and often the reason to choose a protocol instead of another is not explicit.

Many studies also show that some residues far from the active site influence the catalytic mechanism [261,270,271] and therefore the treatment of their flexibility has to be looked upon with care. Here we will evaluate if free / frozen regions of different size influence, in a meaningful way, the activation and reaction QM/MM energies and consequently the free energy of the catalytic mechanism of an enzyme. Note that freezing residues will always have a large effect on the absolute entropy, but here we are focusing on differences between the entropy of different stationary points on the PES of the reaction cycle. For that purpose we explored the first step of the catalytic mechanism of HIV-1 Protease (PR), an enzyme where these effects are expected to be large; it is a small cytosolic-soluble

enzyme, meaning that it has an outer shell of polar residues (to confer solubility) with solvent accessible loops that are close to the active site (due to its small size) followed by two flexible flaps[272,273]. As such, the effect of freezing outer shells of residues will imply changes in residues that establish long-range interactions, and are close to the site of interest, amplifying possible "artifacts" that may arise during the artificial freezing of residues. Additionally, its simplicity, small size and the fact that its mechanism is well established by experimental [274,275] and computational [64,276,277] methods make it ideal for the study.

6.2.2. Protease catalytic mechanism

Two identical chains with 99 residues each compose PR. Its active site has two aspartic catalytic residues. It is well known that one of the two active site aspartates is ionized while the other one is protonated. The first (rate limiting) step of the catalytic mechanism is shown in scheme 1. It is characterized by a nucleophilic attack of a water molecule to the peptide carbonyl of the substrate scissile bond. The nucleophilic water is deprotonated by the negatively charged Asp25 as the attack progresses. The nascent oxyanion is protonated by the neutral $Asp25_B$ still in the same elementary reaction step. This first step leads to the formation of a neutral gem-diol tetrahedral intermediate (**Figure 6.1**). Then, the nitrogen atom of the substrate scissile bond attacks the hydrogen of the Asp25 (protonated during the first step), a proton is transferred from the gem-diol to $Asp25_B$ and the C-N peptide bond breaks to form the reaction products [64,200,213,278-283].

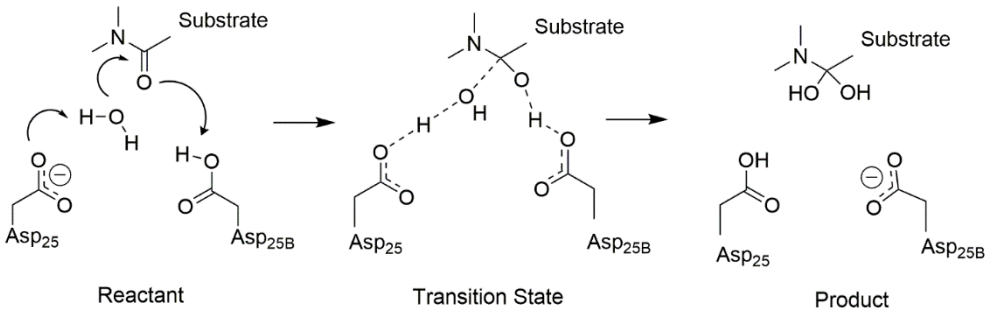


Figure 6.1 The first step of the catalytic mechanism of PR, characterized by a nucleophilic attack of a water molecule on the carbonyl carbon of the substrate scissile bond, forming a tetrahedral intermediate.

To the best of our knowledge there are no works describing in a clear and systematic way the influence of freezing different shells of residues on the activation and reaction QM/MM energies. As such, different protocols are being used, without the desired understanding of the consequences of the underlying choices.

6.3 Methodology

6.3.1 The overall protocol

The computational procedure used in this work started by the modeling of the enzyme–substrate complex using the 4HVP PDB structure [279]. Then a small molecular dynamics (MD) simulation was performed in order to equilibrate the modeled complex. Subsequently, a large MD simulation was run (details in SI) and a representative frame, belonging to the most populated cluster, was used to perform QM/MM calculations with different initial shells of frozen residues.

6.3.2 Model

The PR model was constructed from the 4HVP X-ray structure from the Protein Data Bank (PDB). This initial structure contains the complete enzyme complexed with a substrate-based peptide inhibitor with the sequence Ac-Thr-Ile-Nle-[CH2-NH]-Nle-Gln-Arg.amide. This structure was modeled to a correct substrate with the following sequence: Ac-Thr-Ile-Met-[CO-NH]-Met-Gln-Arg.amide, in the same way as in a previous work by our group [71]. Subsequently, a nucleophilic water molecule was added to the active center and the Asp_{25B} carboxylate was protonated. As mentioned before, it is known that this residue needs to be protonated to initiate the catalytic mechanism of this enzyme. This initial model contains 3232 atoms. We used theGaussView software to model it.

6.3.3 Strategy

With the aim of equilibrating the modeled complex, we performed a first MD simulation without any restriction on the structure. It is well known that a nucleophilic water molecule binds the catalytic center to initiate the catalysis. In this simulation, this water diffused away to the solvent and the catalytic aspartates turned their side chains to each other (the well-known very stable “resting state” of PR). To circumvent this, in a subsequent simulation, we forced the protein to adopt the less abundant catalytic conformation by constraining the distance between the catalytic hydrogen atom of $Asp25_B$ and the carbonyl oxygen atom of the substrate. We used a harmonic potential having an equilibrium length of 1.80 Å between these two atoms and a force constant of 50 $kcal.mol^{-1}Å^{-2}$, using the same protocol as described in a previous work [71].

After the MD simulation (detailed in SI), a snapshot with the nucleophilic water molecule and the catalytic aspartates in a good position to start the reaction, was used in the

QM/MM calculations. The selected structure was present in the most populated cluster of the MD simulation (see **Figure 6.6** (SI) Cluster analysis of the molecular dynamic simulation. We divide all frames (2000) in 10 clusters (from 0 to 9) and they are represented from the most to the less populated one. The yellow point represents the frame that was used in this study). A single shell of water molecules around and inside the protein (~1000 water molecules) was kept in the model. We started by keeping free a shell with a radius of just 4 Å from the residues of the active site. The shell was chosen based on the distance from the atoms of the catalytic residues - $Asp25_A$ (OD1. OD2. CG. CB. HB2 and HB3) and $Asp25_B$ (OD1. OD2. CG. CB. HB3. HB2 and HD2), the nucleophilic water molecule, and the atoms from the substrate scissile bond (CA. HA and O from Met_{201}. and CA. HA. N and O from Met202). These atoms are directly involved in the reaction and they will be designated as “active site” in this work. We repeated the calculation many times, increasing this shell in a gradual way (from 4.00Å free to all free protein) as to increase the number of free atoms in each model.

The activation QM/MM energy and free energy were studied in the same manner for all different models. In all cases, we started with a geometry optimization of the same common reactants structure, changing only the radius of the frozen residues, followed by a scan along the reaction coordinate (the distance between the oxygen of the nucleophilic water molecule and the carbonyl carbon of the scissile peptide bond of the substrate). The geometry that had the highest energy in the linear scan was used as a starting guess to perform a geometry optimization of the transition state. Frequency calculations were conducted to confirm the nature of this structure. Then we performed an intrinsic reaction coordinate calculation (IRC) in order to obtain reaction and product structures in the same relative minimum as the optimized transition state.

6.3.4 ONIOM model and calculations details

Figure 6.2 shows the initial reactants. This structure seems to be in a good orientation to allow the nucleophilic attack of the water molecule on the carbonyl carbon of the substrate and it was obtained after 110 ns of molecular dynamics simulation.

We performed QM/MM calculations, applying an ONIOM scheme as implemented in the Gaussian 09 software package [135]. The system contains a total of 6232 atoms (protein-substrate + water molecules). The QM layer contained 90 atoms (**Figure 6.3**) and the MM layer contained the remaining system. The QM layer contained the two catalytic aspartates (the $Asp25_A$ side chain and the complete $Asp25_B$ residue), the nucleophilic water molecule, seventeen atoms of the substrate, two structural water molecules and some residues around the groups that have an active participation on the reaction

(Ala127, Gly126, Thr125, Gly27, Ala28, and the carbonyl group of Thr_{26}). A list of these atoms is given in table S1. The interaction between the layers was treated with the electrostatic embedding scheme. The QM layer was optimized with the density functional B3LYP [115,206] and 6-31G(d) basis set [125]. Previous works have shown that this functional properly reproduces the geometries and energies for the first step of the PR [64,71,200]. Corrections to the dispersion [123,124] were calculated and added to the final energies.

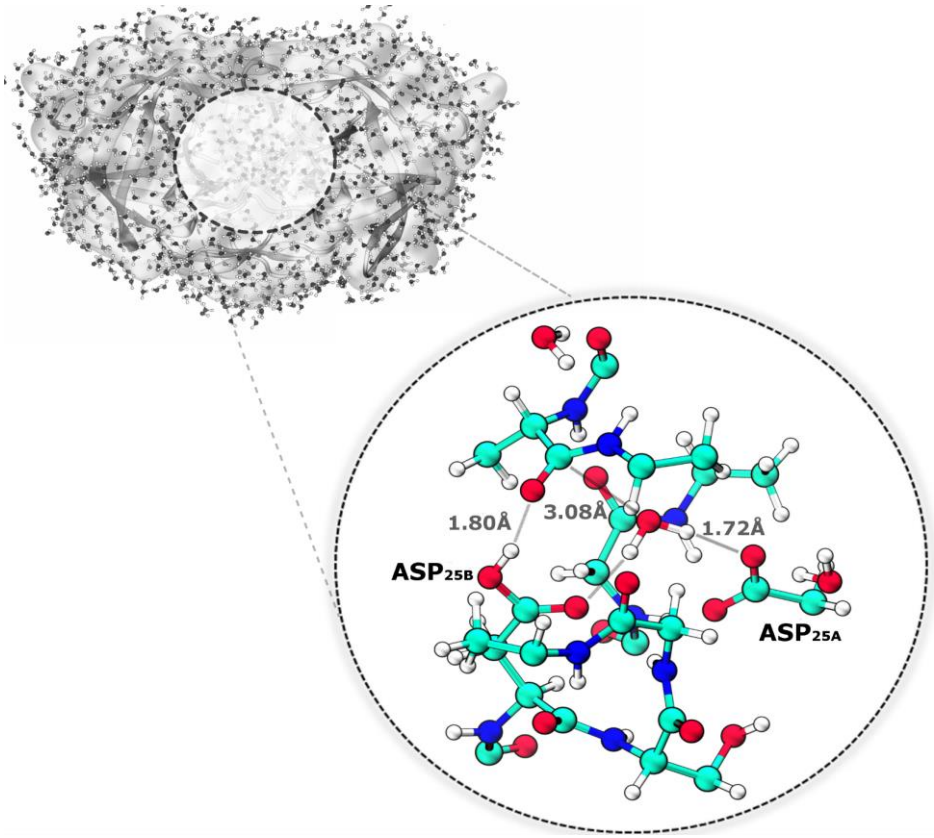


Figure 6.2 Model used in the QM/MM calculations. The high layer (90 atoms) is highlighted in the figure. Relevant distances for the reaction are shown as well.

Some geometry optimizations were performed using the basis set 6-31g(d,p) in order to understand if a better description of the hydrogen atoms would change the results. This was not the case, the optimized stationary points were similar with the 6-31G(d) and the 6-31G(d,p) basis sets. (details in SI). Additionally, some of the stationary points were also optimized with Grimme's D3 [123,124] correction and only small differences were verified (details in SI).

Hydrogen atoms were used as link atoms where QM covalent bonds were truncated. Zero point energies, thermal and entropic corrections were performed at the same theoretical level, using the rigid rotor/harmonic oscillator formalism as implemented in the Gaussian09 software package [144,145]

6.4 Results and Discussion

6.4.1 Dependence of frozen residues in the QM/MM energy profile

We calculated the activation and reaction QM/MM energies and free energies of the first reaction step of PR, keeping free all residues in shells of 4.00 Å, 5.00 Å, 6.00 Å, 7.00 Å, 8.00 Å, 9.00 Å, 10.00 Å, 11.00 Å, 12.50 Å and 15.00 Å radii around the active site.

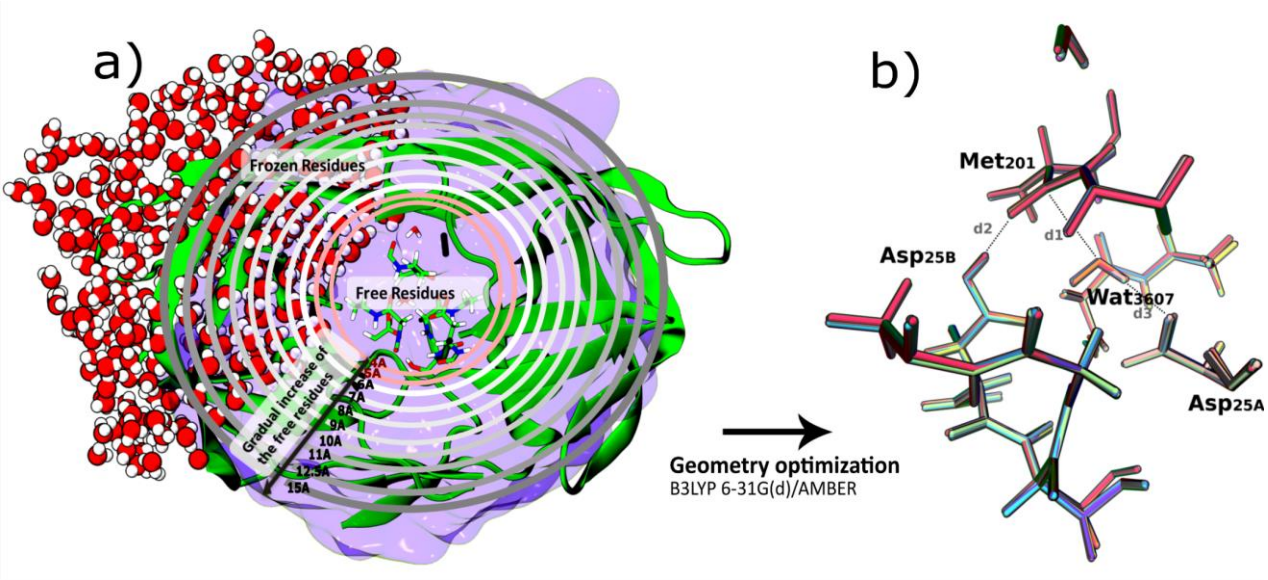


Figure 6.3 a) Schematic representation of the protocol used to study the influence of frozen residues on QM/MM calculations. Eleven models were prepared using the same initial geometry keeping free all residues in the represented shells (from 4.00 to 15.00 Å far from the active site). The last model had all protein/substrate atoms free; for clarity, the solvent is only shown on the top left of the protein; b) Superposition of the geometries obtained after the first geometry optimization for all 11 models (only the high layer is represented). There are no significant differences between them. The relevant distances for the reaction are marked as d1, d2 and d3, and correspond to the distances between Watnuc-O and Met201-C, Asp25B-H and Met201-O, and Wat_{nuc}-H and Asp25A-OD1, respectively: They are very similar between all models (2.78 Å ≤ d1 ≤ 2.81 Å; 1.68 Å ≤ d2 ≤ 1.70 Å; 1.77 Å ≤ d3 ≤ 1.79 Å). These geometries were used as a starting point for the study of PR catalytic mechanism.

The list of free and frozen residues in each model is given in SI. The adopted strategy is schematically represented in **Figure 6.3a**. A last model was prepared keeping all complex enzyme-substrate free (only water molecules were kept frozen).

We compared the structures obtained after the first geometry optimization to be sure that they lie in the same local minimum. The geometries of the QM region obtained for each model are represented in **Figure 6.3b**. All tested models have a similar geometry after optimization. The relevant distances for the reaction (d1, d2 and d3) are particularly similar between all models.

In order to ensure that the differences in the MM region were also small, and that the whole system was in the same local minimum in every structure, the RMSD of the whole protein and the RMSD per residue were calculated for all optimized models (taking the most constrained model as reference). The results showed that the differences between all models were very small (see SI).

Using these obtained geometries, we calculated the activation and reaction QM/MM energies and free energies for the first step of PR. From now on we will focus on the QM/MM energies, as the vibrational contribution to the free energy was calculated in an approximated manner, due to the constraints imposed to the system (even though that did not generate any additional imaginary frequencies).

The obtained barriers (**Table 6.1**) show the impact that the freezing of different layers of residues causes.

Table 6.1 Activation and reaction QM/MM energies (ΔE_{ONIOM}) and free energies (ΔG) (in kcal.mol^{-1}) for enzyme-substrate models with different shells of frozen radius. The differences relative to the free model are shown in the $\Delta\Delta E_{ONIOM}$ columns. Relevant distances (in Å) for the reactions at the TSs are also given. These distances are represented in figure 6.3b. The corresponding distances for the React and Prod states were very similar across all models and were omitted here for simplicity. They can be found together with a larger set of geometry parameters in table S3 (SI).

Free residues	$\Delta E^{\ddagger}_{ONIOM}$	$\Delta E_{R\text{-}ONIOM}$	$\Delta\Delta E^{\ddagger}_{ONIOM}$	$\Delta\Delta E_{R\text{-}ONIOM}$	$\Delta G^{\ddagger}$	ΔG_R	TS imaginary frequency	d1	d2	d3
4.00 Å	22.0	5.0	0.1	4.4	24.1	9.2	217.0i	1.73	1.06	1.62
5.00 Å	24.1	10.2	2.2	0.8	27.0	13.4	51.5i	1.66	1.42	1.62
6.00 Å	22.9	9.9	1.0	0.5	22.8	12.7	421.7i	1.68	1.04	1.41
7.00 Å	22.4	9.9	0.5	0.5	22.9	13.3	442. 5i	1.69	1.04	1.40
8.00 Å	22.2	9.7	0.3	0.3	23.1	13.5	425.6i	1.69	1.04	1.41
9.00 Å	21.8	9.4	0.1	0.0	22.7	12.7	346.9i	1.68	1.04	1.43
10.00 Å	21.8	9.4	0.1	0.0	22.6	12.9	341.2i	1.68	1.04	1.43
11.00 Å	20.5	9.5	1.4	0.1	20.2	12.6	384.5i	1.89	1.48	1.10
12.50 Å	20.3	9.5	1.6	0.1	19.9	12.5	383.1i	1.89	1.48	1.10
15.00 Å	20.3	9.3	1.6	-0.1	20.0	12.4	376.1i	1.89	1.48	1.09
All free	21.9	9.4	-	-	21.9	12.4	376.6i	1.69	1.04	1.42

$\Delta E^{\ddagger}_{ONIOM}$ – ONIOM activation energy; **$\Delta E_{R\text{-}ONIOM}$** – ONIOM reaction energy; **$\Delta\Delta E^{\ddagger}_{ONIOM}$** – Differences in ONIOM activation energy relative to the free model; **$\Delta\Delta E_{R\text{-}ONIOM}$** – Differences in ONIOM reaction energy relative to the free model; **$\Delta G^{\ddagger}$-** Activation free energy; **ΔG_R**- Reaction free energy; **TS imaginary frequency –** Imaginary frequency obtained for each transition state; **d1-** Distance between Wat_{nuc}-O and Met201-C (reaction coordinate); **d2-** Distance between $Asp25_B$-H and Met201-O; **d3-** Distance between Wat_{nuc}-O and Met201-C; **d3-** Wat_{nuc}-H and $Asp25_A$-OD1.

Two large free energy barriers were found in the more constrained models, i.e. $\Delta G^{\ddagger}$ =24.1 kcal.mol^{-1} and $\Delta G^{\ddagger}$ =27.0 kcal.mol^{-1} for shells of 4.00 Å and 5.00 Å of free residues,

respectively. These models are obviously overconstrained. Using the fully free protein as a reference, we studied the differences between this free model and all the other partially frozen models (free radius of 6.00 Å or more). The calculated differences in the barriers ($\Delta\Delta E^{\ddagger}_{ONIOM}$) corresponded, on average, to 0.8 kcal.mol^{-1}, with a maximum difference of 1.6 kcal.mol^{-1}.

The energies of reaction were very similar between all models, with differences in relation to the free model (ΔE_R) that averaged 0.2 kcal.mol^{-1} and a maximum difference of 0.5 kcal.mol^{-1}. The contributions from zero point, thermal and entropic corrections to the system (see SI) were very small and similar between all stationary points of these models, which means that the differences in the entropy between them are very small. It is interesting to see that the products were more influenced with these corrections than the TSs. However, the values were similar between all models.

Table **6.4** shows the geometries of the optimized reactants, transition states and products. We superimposed the structures obtained for the optimized reactants and products of each model. In the optimized reactants (React – **Figure 6.4**) the Wat_{nuc}-carbonyl carbon distance ranged from 2.79 Å to 2.81 Å. A more comprehensive analysis of the geometries is given in SI - table S9. The differences in geometry at the reactant state with different shells of frozen residues were found to be negligible.

The transition state geometries (TS – **Figure 6.4**) show a more interesting difference. The two proton transfers, from Wat_{nuc} to $Asp25_A$ and from $Asp25_B$ to the carbonyl oxygen, take place in all cases, but show different degrees of progress at the TSs of the different models. In most cases the $Asp25_B$ proton is transferred to the carbonyl oxygen before the transfer of the Wat_{nuc} proton to $Asp25_A$ (**Figure 6.4**-TSa)), but in the models with free shells of 5.00 Å, 11.00 Å, 12.50 Å and 15.00 Å, the situation is the opposite, as the TS shows that the substrate carbonyl is still not protonated by $Asp25_B$ and the water molecule is already deprotonated by $Asp25_A$ (**Figure 6.4**-TSb)). The origin of these differences between the TSs will be clarified later on. All the TS geometries were characterized and confirmed as true transition states by vibrational analysis (see **Table 6.1**). In spite of these differences the activation barriers were very similar between studied models (with free radius of 6 Å or more). The geometries of the products (Prod) were similar for all models.

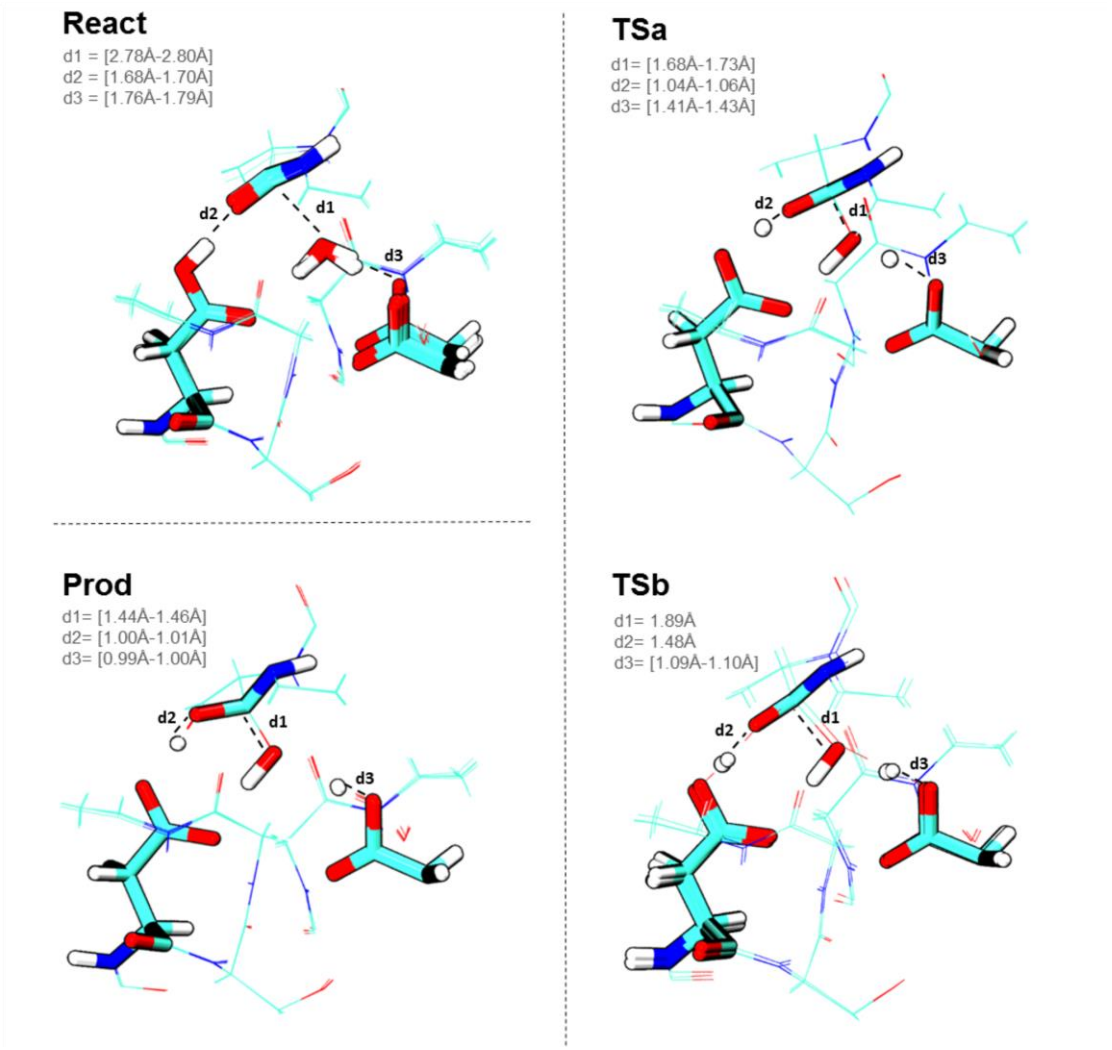


Figure 6.4. Structures of the optimized reactants (React), transition states (TS) and products (Prod) for the first step of the reaction catalyzed by PR. TSa was found in all models except those with 5.00 Å, 11.00 Å, 12.50 Å and 15.00 Å of free residues, where the similar TSb was found. Only the high layer is represented.

In order to understand the contributions of the different regions of the enzyme for the activation and reaction free energies, we divided the ONIOM energies in different components:

$$E_{ONIOM} = E_{QM} + E_{pol} + E_{coul}^{QM/MM} + E_{vdw}^{QM/MM} + E_{prot} \qquad \textbf{Equation 6.1}$$

where E_{QM} corresponds to the energy of the unpolarized QM region (in vacuum), E_{pol} (polarization energy) corresponds to the difference in energy of the QM region in vacuum and within the electrostatic field generated by the MM point charges, $E_{coul}^{QM/MM}$ corresponds to the Coulomb interactions between the QM and MM regions (*i.e.* the interaction between the MM point charges and the QM polarized electronic density, calculated at the QM level), $E_{vdw}^{QM/MM}$ corresponds to the van der Waals interactions

between the QM and MM regions, and the E_{prot} term corresponds to the remaining energy that arises from the MM region (**Table 6.2**) In this analysis, we did not include the overconstrained models (free radius smaller than 6 Å), as they are clearly not recommended for QM/MM studies.

Table 6.2 Activation and reaction energies divided in different components. ΔE_{QM} represents the activation and reaction energies of the unpolarized QM region. ΔE_{pol} represents the polarization energy of the QM region due to the Coulomb interactions with the point charges of the MM region. $\Delta E_{coul}^{QM/MM}$ represents the QM-MM Coulomb interaction energy, calculated with electrostatic embedding. $\Delta E_{vdw}^{QM/MM}$ represents the the QM-MM van der Waals interactions and ΔE_{prot} represents the contribution of the enzyme structure on the ONIOM energy. Grimme's dispersion corrections to the energy were calculated and added to the barriers (last column).

Free residuos	$\Delta E^{\ddagger}$ *ONIOM*	ΔE_R *ONIOM*	$\Delta E^{\ddagger}$ *QM*	ΔE_R, *QM*	$\Delta E^{\ddagger}_{pol}$	ΔE_{R-pol}	$\Delta E^{\ddagger}$ *Coul*	ΔE_{R-Coul}	$\Delta E^{\ddagger}$ *VdW*	ΔE_{R-VdW}	$\Delta E^{\ddagger}$ *prot*	ΔE_{R-prot}	$\Delta E^{\ddagger}$ *ONIOM + D3 Correction*
6.00 Å	22.9	9.9	20.5	18.0	-0.1	0.2	-2.3	-13.5	-2.5	-2.3	7.2	7.4	22.1
7.00 Å	22.4	9.9	20.1	17.5	-0.1	0.2	-2.5	-13.7	-1.5	-1.3	6.4	7.1	22.1
8.00 Å	22.2	9.7	20.1	17.4	-0.1	0.1	-2.3	-13.3	-1.5	-1.3	6.0	6.7	22.0
9.00 Å	21.8	9.4	20.0	17.3	-0.1	0.2	-2.1	-13.1	-1.5	-1.3	5.5	6.4	21.7
10.00 Å	21.8	9.4	20.1	17.2	-0.1	0.1	-2.3	-13.2	-1.2	-1.2	5.3	6.4	21.5
11.00 Å	20.5	9.5	28.0	17.2	0.2	0.1	-10.5	-13.4	-1.2	-0.9	4.0	6.4	20.5
12.50 Å	20.3	9.5	28.0	17.1	0.3	0.2	-10.5	-13.0	-1.3	-1.1	3.9	6.4	21.1
15.00 Å	20.3	9.3	28.0	17.2	0.2	0.2	-10.5	-13.3	-1.5	-1.2	4.0	6.4	20.9
Free	21.9	9.4	20.1	17.3	-0.1	0.1	-2.1	-13.1	-1.3	-1.3	5.2	6.4	21.8

We performed a single point energy calculation using only the QM region at the stationary points of all models, with the same theoretical level. This provided us the QM region energies without the polarizing influence of the MM region (E_{QM}), which is very similar among the several models that react through the same transition structure (TS_a or TS_b), but significantly different (by ~8 kcal·mol^{-1}) when the two transition structures are compared with each other, being the QM barrier of TS_b the higher one.

The polarized QM energy results from the QM energy calculated with the (now frozen) electronic density taken from the electrostatic embedding QM/MM calculations. The polarization energy of the QM region is always quite small, ranging between -0.1 to 0.3 kcal.mol^{-1} in activation energies, and between 0.1 and 0.4 kcal.mol^{-1} in reaction energies.

The contribution of the Coulomb interactions between regions, as calculated through electrostatic embedding, show a similar, albeit inverted, trend in relation to the QM activation energies. In the models that react through TS_a (where $Asp25_B$ proton was transferred to the carbonyl oxygen before the transfer of the Wat_{nuc} proton to $Asp25_A$) these contributions were small (around -2 kcal.mol^{-1} for the activation energy). In turn, in the models that originated TS_b (where the water molecule was deprotonated by $Asp25_A$ before the deprotonation of $Asp25_B$), the Coulomb interactions between the QM region and the MM region amounted to about -10.5 kcal.mol^{-1}, perfectly compensating the higher QM barrier of these models.

The contribution of QM-MM van der Waals energy to the ONIOM energy was calculated subtracting the van der Waals interaction energy of the atoms of the QM region and of the atoms in the MM region to the van der Waals interaction energy of the full system. The contribution of this term on the energetic barriers was, in average, -1.5 kcal.mol^{-1}, with a maximum of -2.5 kcal.mol^{-1}. The differences of this value on the several systems with different frozen regions, (compared with the free model) were very small (in average it was 0.2 kcal.mol^{-1}, with a maximum of 1.2 kcal.mol^{-1}). In general, this term is quite insensitive to the freezing approximation, and thus does not constitute an obstacle to the application of the approximation.

The contribution of the MM region to the barrier, ranged from 5.2 to 7.2 kcal.mol^{-1} in the models that react through TS_a and between 3.9 and 4.0 kcal.mol^{-1} in the models that react through TS_b.

We have then calculated the dispersion corrections for the energy (D3 Grimme's dispersion) in all models and included it in the final results (**Table 6.2** – last column). Interestingly, the differences in activation energy among the models with different frozen regions becomes smaller after adding this correction to the B3LYP energy.

In summary, all activation barriers are very similar in the end, but the systems that evolve through TS_a have a smaller QM barrier and a smaller transition state electrostatic stabilization, and the systems that react through TS_b have a larger QM barrier and a correspondingly larger electrostatic stabilization by the enzyme scaffold.

The freezing of the residues does not affect the energy, or even the partitioning of the energy, even though it can drive the reaction through two similar, quasi-degenerate, different transition structures depending on subtleties of the electrostatic environment.

When we analyzed these energy contributions for the reaction energies, all of them were similar between all models, which is reasonable as the product geometries were similar among them. The Coulomb interactions between layers stabilize the product in about 13 kcal.mol^{-1} (these contributions vary between -10.7 kcal.mol^{-1} and -14.5 kcal.mol^{-1}). The QM-MM van der Waals interaction energy also stabilizes the energy of the reaction (-1.3

kcal.mol^{-1} in average). The contribution of the MM region ranges between 5.1 and 6.4 kcal.mol^{-1}.

6.4.2 Understanding the differences between TS_a and TS_b

It is important to check if the TS differences were related to the freezing protocol or not. First of all, the structural changes were evaluated with a per residue RMSD analysis in each model, having the most constrained model as reference (model 4.00 Å). The results showed that the largest RMSD was around 0.60 Å (only 10 residues had RMSD values above 0.50 Å and most of them are located far from the active site). These residues were highlighted in **Figure 6.5**. The residues colored in green belong to the model that originated TS_b, and the other residues (colored by atom type) belong to the models which originated TS_a. The structural changes were very small and do not justify, per se, the differences between TSs.

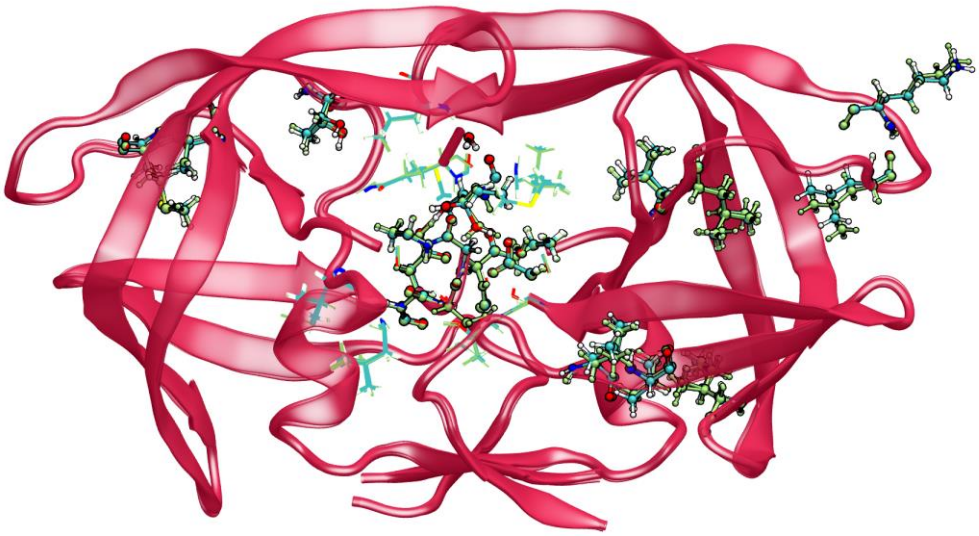


Figure 6.5 Schematic representation of the RMSD per residue for each transition state, having as reference the most constrained one (4.00 Å). The residues with higher RMSD are highlighted. The green residues belong to the models that originate TS_a and the residues colored by atom type belong to models that originate TS_b. The differences between all models are very small.

Subsequently, we tried to understand why the character of the TS changed and why this change introduced large effects on the partition of the total energy among the QM energy and in the QM-MM Coulomb interactions (**Table 6.2**). In TS_b, a hydroxide ion is formed in the active site, and both Asp25 residues were protonated, contrarily to TS_a, where a water molecule making a partial covalent bond to the substrate, two negative Asp25 residues and a partially protonated substrate can be seen. The isolated QM energy was higher for TS_b, as mentioned before. The charge distribution is very different among the two TSs.

However, this same difference in charge distribution will also cause very different interactions with the MM point charges. In fact, the QM region in TS_b was much more stabilized by the Coulomb interaction with the MM layer than the QM region in TS_a. By coincidence, the destabilization in the QM layer in TS_b is compensated by the more favorable QM-MM Coulomb interaction, to a point that both effects cancel out almost completely, making both TSs nearly-degenerate.

To optimize both TSs starting from a same single reactant structure proved to be very hard, due to their geometrical and energetic proximity. We succeeded in doing so in the structure with a free radius of 15.00 Å free (that originally optimized to TS_b). The difference between the TS_a and TS_b was small (TS_b is 1.5 kcal.mol^{-1} lower in energy than TS_a), which means that the two transition states are important in this reaction. Decomposing the energy of TS_a and TS_b for the same enzyme model (see SI)) reveals the same energy partition pattern that was seen for the cases shown in table 2. These results make us believe that the difference between the two TSs is below the level of accuracy of the overall method and protocol. The convergence to one TS or to the other is not dependent on the freezing protocol, but instead in every small methodologic detail (such as the choice of DFT functional or basis set) whose change can slightly affect the energy of the system.

In these models, all water molecules were frozen, even if they were present in the “protein free” region of the system. Therefore, the surface residues kept the same fixed conformation in all systems taken from the MD simulation and are well representative of a solvated protein. These results indicated that freezing a shell of residues of the protein is a very good approach as the changes in activation energy are small. The changes in reaction energy are even smaller, less than 0.6 kcal.mol^{-1}.

6.5 Conclusions

This work focused on the effect of freezing shells of residues during QM/MM calculations of enzyme reaction mechanisms, a very common approximation in the field. We studied the first step of the catalytic mechanism of PR, which is a very good model to study the problem.

The results showed that the frozen residues affect the QM/MM energetic profile of this reaction but to a very small extent. Models with shells of free radius below 6.00 Å showed deviations in the free energies associated with this reaction that, together with the knowledge of the extension of typical atomic rearrangements during chemical reactions, indicated that they were over constrained. With large shells of free residues, the activation energy changes by less than 1.6 kcal.mol^{-1}, depending on the exact radius of free

residues. The structure of the transition state also changes, albeit the changes are not quite relevant to the activation energies of the reaction. The structure of the reactants and products is unaffected by the approximation. The reaction energy was largely insensitive to the shell of frozen residues, showing differences below 0.6 kcal.mol^{-1}.One of the fundamental principles of the study of enzymatic reactions with QM/MM methods is that all the many approximations, of every kind and origin, should provide comparable uncertainties to the calculated energy/free energy, so that the total energy/free energy uncertainty, resulting from the propagation of the individual sources of uncertainty, is not dominated by a single, more severe approximation or assumption. Here we show that the freezing of shells of residues beyond ~6 Å, provides an error (in comparison with a free structure) below 1.6 kcal.mol^{-1} (on average 0.8 kcal.mol^{-1}), which is quite small when compared with errors coming from the choice of the density functional or semiempirical MO/SCC-DFTB method, the basis set used [200], or even the lack of sampling or incomplete sampling [71], which can easily provide errors at least 2-3 times as large.

6.6 Supporting Information

6.6.1 Methodology

6.6.1.2 Details of the molecular dynamics simulation

To perform the MD simulation calculations we started by adding 9960 TIP3P water molecules [105] to the complex protein/modelled substrate in a rectangular box of 88 Å x 67 Å x 71 Å. A minimum of 12 Å were left between any atom of the complex protein-substrate surface and the external molecules of the solvent box. The PME method [284] was used to calculate the Coulombic interactions with the real part truncated at 10 Å. Explicit van der Waals interactions were also truncated at 10 Å. For the simulation without restriction we used the SHAKE algorithm [104] and a time step of 2 fs. When the distance between the side chain proton of Asp_{25B} and the oxygen of the carbonyl carbon of peptide bond was constrained we used a time step of 1 fs. In order to relax the system, removing possible tensions or clashes we started by a three-step minimization of the system using Amber 12 simulation package [285] with parm 99SB force field. First, the water molecules were minimized with the remainder of the system fixed. In these calculations the steepest algorithm for 5000 cycles and conjugated gradient algorithm for the last 5000 steps. Then the hydrogen atoms were minimized, fixing the remainder of the system (steepest descent algorithm to the first 5000 cycles, and conjugate gradient algorithm for the last 5000 steps). Finally, the position of all atoms was minimized (steepest descent algorithm to the first 15000 steps and conjugate gradient algorithm for the last 15000 steps). Starting from the structure obtained after the minimization procedure, we ran molecular dynamics simulations, starting by an initial warm-up of the system from 0 to 300K during a 40 ps long simulation maintaining a constant volume and with periodic boundary conditions (canonical ensemble – NVT). Then we run a MD equilibration on the whole system that run in the isothermal-isobaric ensemble (NPT) with the Langevin thermostat and isotropic position scaling, maintaining the temperature at 300 K and the pressure at 1 bar. Then the production dynamics was run during 200 ns with the same conditions.

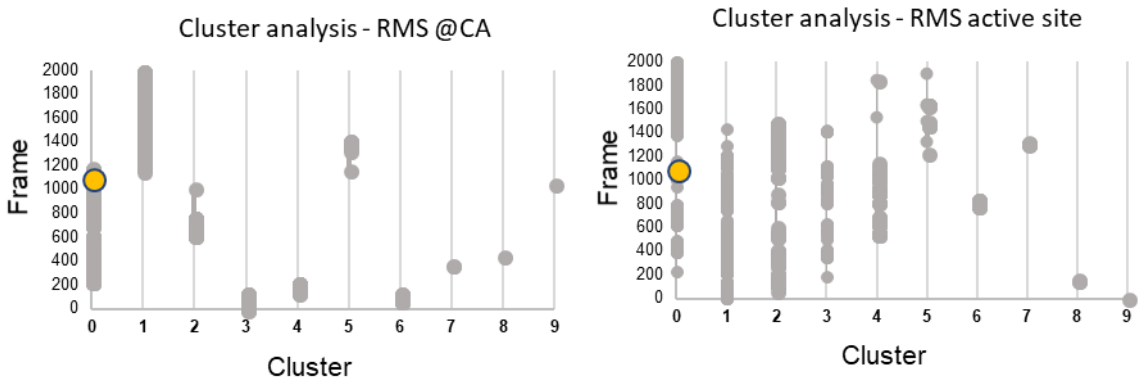


Figure 6.6 (SI) Cluster analysis of the molecular dynamic simulation. We divide all frames (2000) in 10 clusters (from 0 to 9) and they are represented from the most to the less populated one. The yellow point represents the frame that was used in this study

Table 6.3 (SI) List of the atoms in the QM layer.

Resid Number	Resid type	Atoms
25	ASP	CB, HB2, HB3, CG, OD1, OD2
26	THR	C, O
27	GLY	N, H, CA, HA2, HA3, C, O
28	ALA	N, H, CA, HA, CB, HB1, HB2, HB3
123	LEU	C,O
124 (ASP25B)	ASP	N, H, CA, HA, CB, HB2, HB3, CG, OD1, OD2, HD2,C,O
125	THR	N, H, CA, HA, CB, HB, OG1, HG1, C, O
126	GLY	N, H, CA, HA2, HA3, C, O
127	ALA	N, H, CA, HA2, HA3, C,O
200	ILE	C, O
201	MET	N, CA, HA, CB, HB2, HB3, C,O
202	MET	N, H, CA, HA, HB2, HB3
262	WAT	O, H1, H2
3607	WAT	O, H1, H2
9140	WAT	O, H1, H2

Table 6.4 (SI) List of free and frozen atoms in each model

Model	Free Atoms		Frozen Atoms	
Free 4.00 Å	373-453,775-800,1302-1346,1930-2011,2340-2358,2860-2897,3132-3201,3242-3244,4274-4276,5297-5299,5921-5923	Atom Count 373	1-372,454-774,801-1301,1347-1929,2012-2339,2359-2859,2898-3131,3202-3241,3245-4273,4277-5296,5300-5920,5924-6232	Atom Count 5859
Free 5.00 Å	127-164,373-453,775-800,1302-1346,1930-2011,2340-2358,2860-2904,3116-3201,3242-3244,4274-4276,5297-5299,5921-5923,6056-6058	Atom Count 437	1-126,165-372,454-774,801-1301,1347-1929,2012-2339,2359-2859,2905-3115,3202-3241,3245-4273,4277-5296,5300-5920,5924-6055,6059-6232	Atom Count 5795
Free 6.00 Å	127-164,373-453,768-800,1302-1370,1404-1422,1930-2011,2333-2358,2816-2845,2860-2904,2962-2980,3116-3226,3242-3244,3416-3418,4274-4276,5921-5923	Atom Count 565	1-126,165-372,454-767,801-1301,1371-1403,1423-1929,2012-2332,2359-2815,2846-2859,2905-2961,2981-3115,3227-3241,3245-3415,3419-4273,4277-5920,5924-6232	Atom Count 5667

Model	Free Atoms		Frozen Atoms	
Free 7.00 Å	127-164,373-507,768-807,1288-1370,1404-1422,1708-1721,1920-2065,2307-2358,2816-2928,2962-2980,3116-3226,3242-3244,4274-4276,5921-5923	Atom Count 779	1-126,165-372,508-767,808-1287,1371-1403,1423-1707,1722-1919,2066-2306,2359-2815,2929-2961,2981-3115,3227-3241,3245-4273,4277-5920,5924-6232	Atom Count 5453
Free 8.00 Å	67-85,127-164,363-507,749-814,1244-1257,1272-1370,1404-1422,1684-1740,1920-2065,2307-2358,2802-2928,2962-2980,3062-3080,3116-3226,3242-3244,4274-4276,5921-5923	Atom Count 940	1-66,86-126,165-362,508-748,815-1243,1258-1271,1371-1403,1423-1683,1741-1919,2066-2306,2359-2801,2929-2961,2981-3061,3081-3115,3227-3241,3245-4273,4277-5920,5924-6232	Atom Count 5292
Free 9.00 Å	67-85,110-199,363-507,749-814,835-853,1244-1257,1272-1422,1504-1522,1624-1642,1684-1740,1920-2084,2307-2372,2393-2411,2802-2980,3062-3080,3116-3226,3242-3244,3785-3787,4031-4033,4274-4276,5795-5797,5849-5851,5885-5887,5921-5923	Atom Count 1182	1-66,86-109,200-362,508-748,815-834,854-1243,1258-1271,1423-1503,1523-1623,1643-1683,1741-1919,2085-2306,2373-2392,2412-2801,2981-3061,3081-3115,3227-3241,3245-3784,3788-4030,4034-4273,4277-5794,5798-5848,5852-5884,5888-5920,5924-6232	Atom Count 5050
Free 10.00 Å	67-85,110-199,363-507,749-853,1188-1206,1244-1422,1504-1522,1624-1642,1684-1756,1920-2084,2307-2372,2393-2411,2802-2980,3012-3030,3062-3080,3116-3226,3242-3244,4274-4276,5921-5923	Atom Count 1255	1-66,86-109,200-362,508-748,854-1187,1207-1243,1423-1503,1523-1623,1643-1683,1757-1919,2085-2306,2373-2392,2412-2801,2981-3011,3031-3061,3081-3115,3227-3241,3245-4273,4277-5920,5924-6232	Atom Count 4977
Free 11.00 Å	1065,1188-1206,1244-1436,1454-1472,1480-1489,1504-1522,1591-1609,1624-1642,1684-1756,1771-1789,1905-2084,2307-2411,2746-2764,2802-2980,3012-3030,3038-3047,3062-3080,3095-3226,3242-3244,4274-4276,5921-5923	Atom Count 1497	046,1066-1187,1207-1243,1437-1453,1473-1479,1490-1503,1523-1590,1610-1623,1643-1683,1757-1770,1790-1904,2085-2306,2412-2745,2765-2801,2981-3011,3031-3037,3048-3061,3081-3094,3227-3241,3245-4273,4277-5920,5924-6232	Atom Count 4735
Free 12.50 Å	34-52,67-199,214-232,348-541,732-853,876-891,1047-1065,1158-1206,1230-1436,1454-1472,1480-1489,1504-1522,1537-1557,1591-1609,1624-1642,1667-1789,1814-1832,1883-2099,2307-2411,2434-2449,2605-2623,2716-2764,2788-3030,3038-3047,3062-3080,3095-3226,3242-3244,4274-4276	Atom Count 1843	1-33,53-66,200-213,233-347,542-731,854-875,892-1046,1066-1157,1207-1229,1437-1453,1473-1479,1490-1503,1523-1536,1558-1590,1610-1623,1643-1666,1790-1813,1833-1882,2100-2306,2412-2433,2450-2604,2624-2715,2765-2787,3031-3037,3048-3061,3081-3094,3227-3241,3245-4273,4277-6232	Atom Count 4389
Free 15.00 Å	1-232,257-275,326-556,710-891,1016-1031,1047-1075,1122-1131,1151-1573,1591-1789,1814-1832,1883-2099,2115-2131,2268-2449,2574-2589,2605-2623,2709-3226,3242-3244,4274-4276,5921-5923	Atom Count 2338	233-256,276-325,557-709,892-1015,1032-1046,1076-1121,1132-1150,1574-1590,1790-1813,1833-1882,2100-2114,2132-2267,2450-2573,2590-2604,2624-2708,3227-3241,3245-4273,4277-5920,5924-6232	Atom Count 3894
All Free	1-3226,3242-3244,4274-4276,5921-5923 (Protein + Substrate+ 3 Water molecules)	Atom Count 3235	3227-3241,3245-4273,4277-5920,5924-62 (Water molecules)	Atom Count 2997

6.6.2 Results

6.6.2.1 Comparison between all models after a geometry optimization

Table 6.5 RMSD (Å) of all models after a geometry optimization, with ONIOM(B3LYP/6-31G(d):ff99SB) level of theory, relative to the most constrained one. The values were very similar between all models, which means that there were not significant changes in the structures.

Model	RMSD (Å)
4.00 Å	0.00
5.00 Å	0.07
6.00 Å	0.09
7.00 Å	0.12
8.00 Å	0.14
9.00 Å	0.16
10.00 Å	0.17
11.00 Å	0.20
12.50 Å	0.23
15.00 Å	0.26
Free	0.33

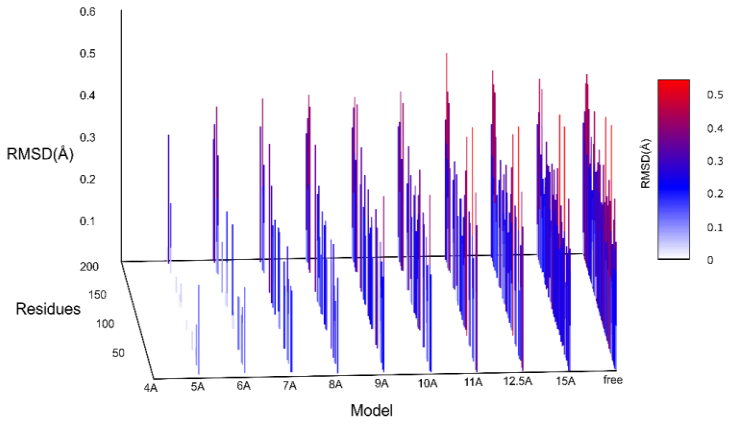


Figure 6.7 (SI) Schematic representation of the RMSD by residue for each model after a geometry optimization, having as reference the most constrained one (4.00 Å). The differences between all models were small (the highest RMSD value was near 0.5Å). The highest values were red represented.

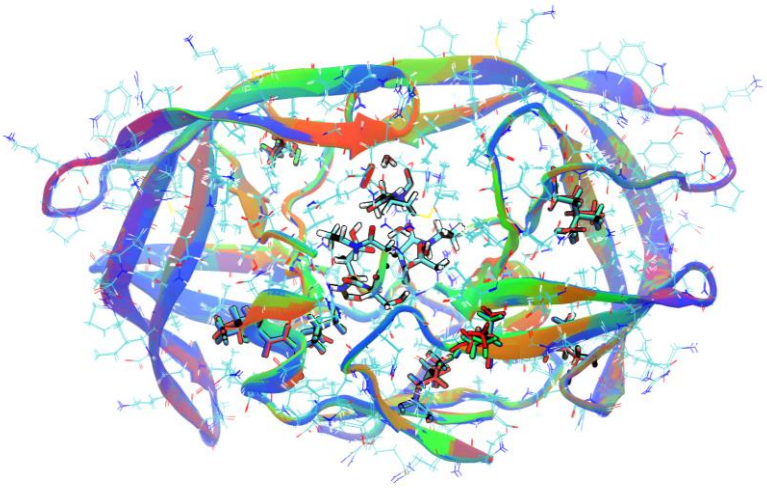


Figure 6.8 (SI) The optimized models are superimposed and the residues with higher RMSD are highlighted. The differences between all models are very small.

6.6.2.2 Comparison between all transition states

Table 6.6 (SI) RMSD (Å) of all optimized states relative to the most constrained model. ONIOM(B3LYP/6-31G(d):ff99SB) level of theory was used.

Model	RMSD (Å)
4.00Å	0.00
5.00Å	0.04
6.00Å	0.07
7.00Å	0.11
8.00Å	0.11
9.00Å	0.13
10.00Å	0.14
11.00Å	0.17
12.50Å	0.18
15.00Å	0.22
Free	0.27

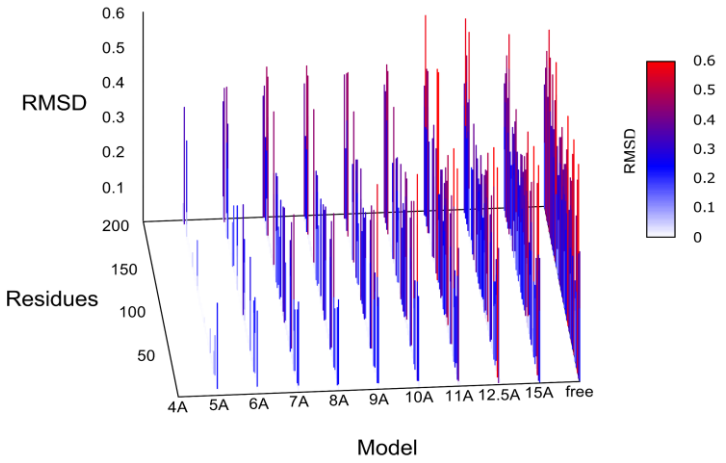


Figure 6.9 (SI)Schematic representation of the RMSD by residue for each optimized transition state, having as reference the most constrained one (4.00Å).

6.6.2.3 Results with a different basis set (6-31G(d,p))

Table 6.7 (SI) Absolute and relative energies for the reactants and transition states optimized with different basis sets. The test was performed only for three representative models.

		Energy			
		ONIOM(B3LYP/6-31G(d):ff99SB)		ONIOM(B3LYP/6-31G(d,p):ff99SB)	
Free residues	Stationary points	E_{ONIOM} (Hartree)	E_{ONIOM} (kcal.mol^{-1})	E_{ONIOM} (Hartree)	E_{ONIOM} (kcal.mol^{-1})
6.00 Å	R	-2643.132	**0.0**	-2643.251	**0.0**
	TS	-2643.096	**22.9**	-2643.217	**21.3**
15.00 Å	R	-2645.384	**0.0**	-2645.503	**0.0**
	TS	-2645.352	**20.3**	-2645.472	**19.3**
All Free	R	-2646.410	**0.0**	-2646.529	**0.0**
	TS	-2646.375	**22.0**	-2646.496	**20.8**

Table 6.8 (SI) Main distances for the reactants and transition state points optimized with different basis sets. The test was performed only for three representative models.

		Distances / Å					
		ONIOM(B3LYP/6-31G(d):ff99SB)			ONIOM(B3LYP/6-31G(d,p):ff99SB)		
Free residues	Stationary points	d1	d2	d3	d1	d2	d3
6.00 Å	R	2.81	1.69	1.78	2.77	1.66	1.77
	TS	1.68	1.04	1.41	1.71	1.06	1.40
15.00 A	R	2.77	1.69	1.78	2.77	1.66	1.77
	TS	1.89	1.48	1.09	1.86	1.40	1.10
All Free	R	2.78	1.69	1.78	2.80	1.65	1.78
	TS	1.68	1.04	1.42	1.71	1.06	1.41

6.6.2.4 Results with correction for dispersion

Table 6.9 (SI) Main distances for the reactants and transition state points optimized with and without Grimme D3 correction. The test was performed only for two representative models.

		Energy			
		ONIOM(B3LYP/6-31G(d):ff99SB)		ONIOM(B3LYP/6-31G(d):ff99SB) + D3 Grimme's correction	
Free residues	Stationary points	E_{ONIOM} (Hartree)	E_{ONIOM} (kcal.mol^{-1})	E_{ONIOM} (Hartree)	E_{ONIOM} (kcal.mol^{-1})
6.00 Å	R	-2643.132	**0.0**	-2643.285	**0.0**
	TS	-2643.096	**22.9**	-2643.253	**20.2**
	P	-2643.117	**9.9**	-2643.269	**9.9**
15.00 Å	R	-2645.384	**0.0**	-2645.538	**0.00**
	TS	-2645.352	**20.3**	-2645.507	**19.3**
	P	-2645.369	**9.3**	-2645.523	**9.18**

Table 6.10 (SI) Main distances for the stationary points optimized with and without Grimme D3 correction. The test was performed only for two representative models

		Distances / Å					
		ONIOM(B3LYP/6-31G(d):ff99SB)			ONIOM(B3LYP/6-31G(d):ff99SB) + D3 Grimme's correction		
Free residues	Stationary points	d1	d2	d3	d1	d2	d3
6.00 Å	R	2.81	1.69	1.78	2.70	1.64	1.76
	TS	1.68	1.04	1.41	1.68	1.04	1.40
	P	1.46	1.00	1.00	1.47	1.00	1.00
15.00 A	R	2.77	1.69	1.78	2.67	1.65	1.76
	TS	1.89	1.48	1.09	1.84	1.41	1.09
	P	1.46	1.00	1.00	1.46	1.01	1.00

Table 6.11 (SI) Key distances in the optimized reactants, transition states and products from models with a different shell of free / frozen residues (The results are represented in Å). ONIOM(B3LYP/6-31G(d):ff99SB) level of theory was used

		Distance	Shell of free residues around the active site									
			4.00 Å	5.00 Å	6.00 Å	7.00 Å	8.00 Å	9.00 Å	10.00 Å	11.00 Å	12.50Å	15.00 Å	Free
R	*d1*	Wat_{nuc}-O … Met_{201}-C	2.81	2.82	2.81	2.81	2.80	2.79	2.78	2.78	2.78	2.79	2.78
	d2	Asp_{25B}-H_{d2} … Met_{201}-O	1.68	1.69	1.69	1.68	1.68	1.70	1.69	1.69	1.69	1.69	1.69
	d3	Asp_{25A}-O_{d2} …Wat_{nuc}-H_1	1.78	1.76	1.79	1.78	1.78	1.78	1.78	1.79	1.78	1.78	1.78
	d4	Asp_{25B}-O_{d1} … Wat_{nuc}-H_2	1.98	1.98	1.93	1.94	1.94	1.94	1.93	1.93	1.94	1.95	1.94
	d5	Wat_{nuc}-H_1 … Asp_{25A}-O_{d1}	2.66	2.66	2.71	2.67	2.67	2.66	2.66	2.65	2.65	2.66	2.66
	d6	Asp_{25B}-O_{d1} … Gly_{126}-H	2.59	2.56	2.63	2.62	2.60	2.59	2.62	2.63	2.61	2.60	2.59
	d7	Asp_{25A}-O_{d1} … Trn_{125}-H_{g1}	3.02	3.02	3.43	3.38	3.41	3.45	3.44	3.47	3.47	3.47	3.47
	d8	Wat_{nuc}-O … Met_{201}-H	2.51	2.49	2.54	2.52	2.48	2.44	2.44	2.44	2.45	2.45	2.45
	d9	Wat_{9140}-H_2 … Asp_{25}-O_{d2}	1.80	1.81	1.81	1.82	1.82	1.84	1.83	1.83	1.83	1.83	1.83
		Distance	**Shell of free residues around the active site**										
			4.00 Å	**5.00 Å**	**6.00 Å**	**7.00 Å**	**8.00 Å**	**9.00 Å**	**10.00 Å**	**11.00 Å**	**12.50 Å**	**15.00 Å**	**Free**
TS	*d1*	Wat_{nuc}-O … Met_{201}-C	1.73	1.66	1.68	1.69	1.69	1.68	1.68	1.89	1.89	1.89	1.69
	d2	Asp_{25B}-H_{d2} … Met_{201}-O	1.06	1.42	1.04	1.04	1.04	1.04	1.04	1.48	1.48	1.48	1.04
	d3	Asp_{25A}-O_{d2} …Wat_{nuc}-H_1	1.62	1.62	1.41	1.41	1.41	1.43	1.43	1.10	1.10	1.09	1.42
	d4	Asp_{25B}-O_{d1} … Wat_{nuc}-H_2	1.59	1.73	1.61	1.62	1.62	1.60	1.60	2.04	2.04	2.04	1.62
	d5	Wat_{nuc}-H_1 … Asp_{25A}-O_{d1}	2.65	2.71	2.55	2.53	2.53	2.53	2.53	2.41	2.41	2.41	2.53
	d6	Asp_{25B}-O_{d1} … Gly_{126}-H	2.39	2.48	2.31	2.30	2.29	2.34	2.32	2.16	2.16	2.15	2.28
	d7	Asp_{25A}-O_{d1} … Trn_{125}-H_{g1}	3.30	3.33	3.78	3.71	3.75	3.75	3.77	3.69	3.71	3.70	3.80
	d8	Wat_{nuc}-O … Met_{201}-H	2.63	2.61	2.68	2.67	2.67	2.67	2.67	2.43	2.43	2.43	2.66
	d9	Wat_{9140}-H_2 … Asp_{25}-O_{d2}	1.91	1.93	2.10	2.11	2.11	2.14	2.13	2.28	2.27	2.23	2.13

(Cont.) Key distances in the optimized reactants, transition states and products from models with a different shell of free / frozen residues (The results are represented in Å). ONIOM(B3LYP/6-31G(d):ff99SB) level of theory was used

		Distance	Shell of free residues around the active site										
			4.00 Å	5.00 Å	6.00 Å	7.00 Å	8.00 Å	9.00 Å	10.00 Å	11.00 Å	12.50 Å	15.00 Å	Free
P	*d1*	Wat_{nuc}-O … Met_{201}-C	1.44	1.46	1.46	1.46	1.46	1.46	1.46	1.46	1.46	1.46	1.46
	d2	Asp_{25B}-H_{d2} … Met_{201}-O	1.00	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
	d3	Asp_{25A}-O_{d2} …Wat_{nuc}-H_1	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	d4	Asp_{25B}-O_{d1} … Wat_{nuc}-H_2	1.73	1.76	1.75	1.75	1.75	1.75	1.75	1.75	1.75	1.76	1.76
	d5	Wat_{nuc}-H_1 … Asp_{25A}-O_{d1}	2.33	2.33	2.34	2.33	2.33	2.33	2.33	2.33	2.33	2.33	2.33
	d6	Asp_{25B}-O_{d1} … Gly_{126}-H	2.16	2.17	2.14	2.14	2.13	2.14	2.14	2.12	2.13	2.12	2.12
	d7	Asp_{25A}-O_{d1} … Trn_{125}-H_{g1}	3.15	3.15	3.65	3.59	3.62	3.75	3.66	3.69	3.69	3.69	3.69
	d8	Wat_{nuc}-O … Met_{201}-H	3.73	2.70	2.73	2.73	2.72	2.72	2.72	2.72	2.72	2.72	2.72
	d9	Wat_{9140}-H_2 … Asp_{25}-O_{d2}	2.11	2.13	2.21	2.21	2.22	2.27	2.26	1.36	2.27	2.25	2.24

Table 6.12 (SI) Imaginary frequencies of the transitions states for each model; ONIOM energetic barriers and the contributions from zero point energy (ZPE), thermal and entropic corrections (values in kcal.mol-1). ONIOM(B3LYP/6-31G(d):ff99SB) level of theory was used.

Free residues	TS imaginary frequency	$\Delta E^{\ddagger}_{ONIOM}$	ZPE + Thermal Correction - $T\Delta S^{\ddagger}$	ΔE^{R}_{ONIOM}	ZPE +Thermal Correction - $T\Delta S_R$
4.00 Å	217.0i	22.0	2.1	5.0	4.2
5.00 Å	51.5i	24.1	2.7	10.2	4.0
6.00 Å	421.7i	22.9	-0.1	9.9	2.7
7.00 Å	442. 5i	22.4	0.5	9.9	3.4
8.00 Å	425.6i	22.2	0.9	9.7	3.7
9.00 Å	346.9i	21.8	-0.1	9.4	3.3
10.00 Å	341.2i	21.8	0.8	9.4	3.5
11.00 Å	384.5i	20.5	-0.3	9.5	3.1
12.50 Å	383.1i	20.3	-0.4	9.5	3.0
15.00 Å	376.1i	20.3	-0.2	9.3	3.1
Free	376.6i	21.9	0.0	9.4	3.0

Table 6.13 (SI) Energy differences between the geometry of TSa and TSb for the model 15Å. The partition of the energies shows that the barrier in TSa comes mostly from the QM region; in TSb the barrier is quite high in the QM region but smoothed down by a very favorable Coulomb interaction with the MM layer. ONIOM(B3LYP/6-31G(d):ff99SB) level of theory was used.

Free	TS	$\Delta E^{\ddagger}_{ONIOM}$	$\Delta E^{\ddagger}_{QM}$	$\Delta E^{\ddagger}_{pol}$	$\Delta E^{\ddagger}_{Coul}$	$\Delta E^{\ddagger}_{VdW}$	$\Delta E^{\ddagger}_{prot}$
15.00 Å	**b**	**20.3**	**28.0**	**0.2**	**-10.5**	**-1.5**	**4.0**
	a	21.8	20.0	-0.03	-2.1	-1.2	5.1

CHAPTER 7. A buried water molecule influences reactivity in alpha-amylase on a sub-nanosecond timescale

Diogo Santos-Martins, Ana Rita Calixto, Pedro Alexandrino Fernandes and Maria João Ramos

UCIBIO, REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal,

In the following work, the first step of the catalytic mechanism of alpha-amylase was explored starting from different initial structures taken from a molecular dynamics simulation, to account for the influence of enzyme flexibility on this enzyme.

The results showed that the activation barrier of this mechanism is dependent on the position and orientation of a buried, non-catalytic, water molecule.

Regarding the contributions to the paper, Diogo Santos Martins performed the first part of the calculations and wrote the first draft manuscript. Ana Rita Calixto performed the final calculations, the structural analysis and wrote some parts of the manuscript which was revised through contribution of all authors. This work was published in the *ACS Catalysis*, and the following content corresponds to an integral transcription of the published version

Santos-Martins, D.; **Calixto, A. R**.; Fernandes, P. A. and Ramos M. J., ***Water Controls Reactivity in Alpha-amylase on a Subnanosecond Timescale***. ACS Catalysis, 2018, 8, 4055-4063 **DOI: 10.1021/acscatal.7b04400**

7.1 Abstract

The subset of catalytically competent conformations can be significantly small in comparison with the full conformational landscape of enzyme-substrate complexes. In some enzymes, the probability of finding a reactive conformation can account for up to 4 kcal/mol of activation barrier, even when the substrate remains tightly bound. In this study, we sampled conformations of human pancreatic alpha-amylase with bound substrate in a molecular dynamics (MD) simulation of over 100 ns, and calculated energy profiles along the reaction coordinate. We found that reactive states require a hydrogen bond between a buried water molecule and E233, which is the general acid in the glycolysis mechanism. The effect of this single, non-reactive, intermolecular interaction is as important as the correct positioning and orientation of the reacting residues to achieve a competent energy barrier. This hydrogen bond increases the acidity of E233, facilitating proton transfer to the glycosidic oxygen. In the MD simulation, this required hydrogen bond was observed in more than half of the microstates, indicating that human pancreatic alpha-amylase is efficient at maintaining this important interaction in the reactants state. Furthermore, this hydrogen bond formed and vanished on a sub-nanosecond time scale. Interactions between the reacting groups, also change at this timescale. All these changes led to instantaneous activation energy oscillations from 9.3 kcal/mol to 28.3 kcal/mol at a much smaller timescale than turnover rate. These results are in agreement with observed kinetics being determined by a few, transient, conformations that require low energy barriers.

Keywords
Enzymatic Catalysis, Reactive Enzyme Conformations, Transition State Stabilization, Alpha-amylase, Carbohydrates.

7.2 Introduction

Chemical reactions only occur if the reacting molecules are in close proximity and in a suitable orientation, otherwise the electronic rearrangements associated with the reaction do not take place. Therefore, observed free energy barriers are proportional to the probability of finding the reactants in suitable (reactive) conformations, no matter how low this probability may be.

In fact, there are extreme examples of this effect, such as the isomerization of chorismate to prephenate in which reactive conformers correspond to only 0.0001% of the conformational space [26,286,287], contributing 8.4 kcal/mol to the free energy of activation [55]. Still associated to this latter reaction, it is important to state that the correspondent free energy penalty is not linked to changes in the electronic structure of chorismate, but instead with the probability of finding chorismate in a suitable geometry for the reaction to take place. Based on extensive studies of the aforementioned reaction, the activation free energy $\Delta G^{\ddagger}$ was decomposed into chemical and nonchemical [26,55]:

$$\Delta G^{\ddagger} = \Delta G_{NAC} + \Delta G_{Chem} \qquad \textbf{Equation 7.1}$$

where ΔG_{NAC} is the free energy difference between reactive substrate conformations (NAC stands for near attack conformation) and the full conformational space of the reactants (the ground state), and ΔG_{Chem} is the free energy of activation associated with the chemical transformation of a NAC to a transition state (TS). The chemical component is primarily associated with the changes in electronic energy, zero-point energy and its stabilization by the surrounding environment — its value reflects the amount of kinetic energy necessary to overcome the TS. On the other hand, the non-chemical component ΔG_{NAC} reflects the probability of finding the reagents in a reactive conformation, which is 8.4 kcal/mol for the isomerization of chorismate.

Here, we will define reactive enzyme conformations as the subset of microstates that obey geometrical criteria for reactions to take place with low chemical barriers. Such geometrical criteria can be applied to enzyme-substrate complexes and to uncatalyzed reactions independently from each other, which is useful because catalyzed and uncatalyzed reactions often follow different mechanisms. Reactive enzyme conformations can also apply strictly to the conformation of the substrate or, more broadly, to the conformation of the substrate and active site (or, less frequently, to the complete enzyme and solvent), as its residues do act chemically/electrostatically over the substrate and need to obey strict geometrical constraints to do so in an efficient manner, despite the overall conservation of the catalyst at the end of the enzymatic cycle. It is worth noting

that the geometrical definitions of a reactive enzyme conformation are necessarily subjective, and the meaning of the energy values associated with these conformations are tied to the quality of the underlying geometrical definitions. In enzymes this energy may be surprisingly large in view of the conformational confinement of substrates inside active sites, which are expected to maintain required interactions for transition state stabilization (such as hydrogen bonds) consistently throughout the existence of a reactive enzyme-substrate complex. Molecular dynamics (MD) simulations estimate an energy up to 4.6 kcal/mol for the formation of reactive conformations in HIV-1 protease [17] and a similar value of 4 kcal/mol was calculated for barnase [288]. By examining why an enzyme design failed, Ruscio et. al concluded that enzyme design needs to verify a reactive conformational condition using a dynamical approach [289]. These studies indicate that enzyme-substrate complexes navigate a complex conformational landscape where critical interaction for catalysis have a remote chance of occurrence, at least in some enzymes. In computational studies this problem is often unseen as many authors who use protocols based in cluster models or QM/MM geometry optimizations with high QM theoretical levels already start from a catalytically conformation, and easily get productive PES, providing accurate calculations of ΔG_{Chem}.

The strong dependence of reactivity on the enzyme conformation is well documented by a large body of studies, in which activation barriers were calculated for several different conformations of the same enzyme. These studies typically employ QM/MM methodologies with adiabatic mapping protocols to compute activation energies/free energies, reliably informing about the reactivity of each conformation. For example, a conformational change in ketosteroid isomerase raised the activation energy by *circa* 20 kcal/mol [290], variations up to 17 kcal/mol were found in P450 catalyzed reactions [291] and in fatty acid amide hydrolase the range of activation barriers was 11 kcal/mol [292]. Many other studies found significant variations of activation barriers for varying enzyme-substrate conformations [64,71,212,293-305].

In our previous study on HIV-1 protease [71], interactions of the nucleophilic water in the active site and the alignment of charge transfer within reactive groups with the electrostatic potential generated by the whole enzyme explained most of the observed fluctuations in the barriers. Interestingly, such fluctuations occur on the nanosecond timescale, while the turnover rate is on the second-time scale. This means that even enzymes without dynamic disorder [306,307], which appear to have a constant rate throughout many cycles, should experience ‘instantaneous disorder”. Instantaneous disorder amounts to the fluctuations of activation energy on a timescale order of magnitude faster than turnover rate. In view of the fact that chemical reactions occur on a femtosecond timescale [159,308], chemical

reactions probably take place at specific conformations with low barriers. The observed rate constant depends mostly on both the frequency of low barriers and their magnitude.
In this work, we identify structural features of reactive conformations of on human pancreatic α-amylase (HPA). We performed MD and ONIOM QM/MM studies on HPA, addressing the first half reaction of the catalytic cycle, focusing on the relationship between activation free energies calculated within the single-conformation model and interactions occurring in the active site for each given single-conformation. HPA is a good case study because the glycosylation mechanism is well established [235] and is also relevant to the glycoside hydrolase superfamily, a very large class of enzymes implied in a variety of diseases [309].
Our results highlight the surprising importance of a buried water in determining the reactivity in HPA. The orientation of this water was as important as the orientation of the catalytic residues. Therefore, we can think of this water a structural determinant of reactivity, and as part of the catalytic machinery of HPA.

7.3 Methods

Our work consisted of two main stages: (i) conformational sampling of the reactants state of human alpha-amylase over 100 ns of molecular dynamics (MD) simulation and (ii) calculation of activation energies and free energies, within the single-conformation model, starting from different selected microstates of the MD trajectory.

7.3.1 Molecular Dynamics

7.3.3.1 System Details

HPA was simulated by molecular dynamics (MD) with substrate maltopentaose (G5). Initial coordinates were retrieved from PDB [136] structure '1CPU', which was co-crystallized with an acarbose-like inhibitor with five monosaccharide rings [310]. Modeling G5 from this inhibitor involved two simple modifications: changing the N-glycosidic bond to an O-glycosidic bond and replacing a C=C double bond by a C-O single bond in the ring adjacent to the N-glycosidic bond of the inhibitor. The binding mode of G5 was such that the catalytic machinery of HPA was positioned to cleave the glycosidic bond between the third and fourth glucose units, producing maltotriose (G3) and maltose (G2). In the X-ray structure N461 was glycosylated, which was achieved by removing the carbohydrate. All titratable groups were simulated in their standard protonation states (pH=7) except for

E233 which was simulated in the neutral state, as required for the reaction to take place. E233 works as the general acid and donates a proton to the scissile glycosidic bond [235].

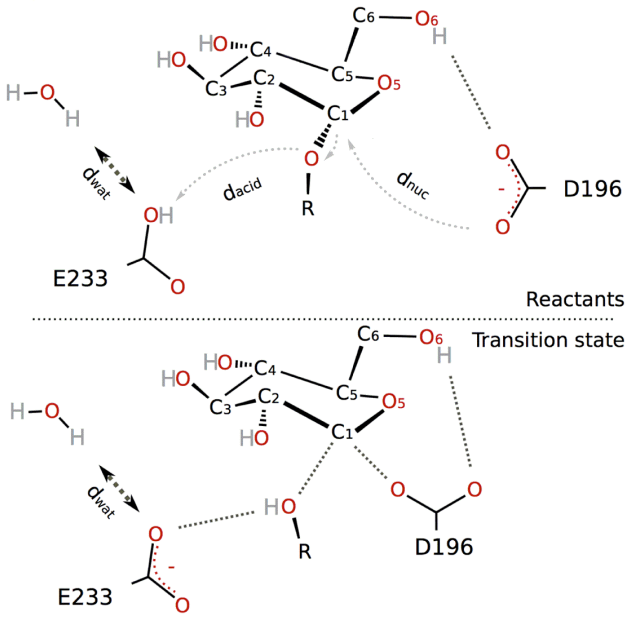


Figure 7.1 Reactants and transition state of the glycolysis step. Important distances are defined: d_{wat}, established between a water hydrogen and the protonated oxygen of E233, d_{acid} between the acidic hydrogen of E233 and the glycosidic oxygen, d_{nuc} between the C1 and a carboxylate oxygen of D196.

The hydrogen bond corresponding to d_{acid} in **Figure 7.1** was restrained at 2.0 Å with a harmonic potential with a force constant of 50 kcal· mol^{-1} ·Å^{-2} in order to enrich the sampling in catalytically competent conformations. The system was solvated in a periodic box filled with TIP3P water molecules such that at least 12 Å exist between the protein or G5 and the edges of the box. Sodium counterions were added to neutralize the charge of the system.

7.3.2.1 Simulation

Simulations were carried out using AMBER 12 [103]. The ff99SB forcefield [311] was used for HPA, the TIP3P model [105] for water molecules and GLYCAM_06h [312] parameters for G5. The leap program in Antechamber [313] was used to assign parameters. Particle Mesh Ewald (PME) [284,314,315] was used for electrostatics, with the real part truncated at 10 Å. The integration time step was 2 fs, using the Shake algorithm [104]. Langevin dynamics was used with a collision frequency of 1 ps, using the NPT ensemble at 1 bar and a pressure relaxation time of 2 ps, at 300 K. Microstates were recorded every 100 ps. The total simulation time for this production run was 109 ns. Waters and counterions were

previously equilibrated by a 10 ns NVT ensemble run, maintaining all protein and G5 atoms fixed at their crystallographic coordinates. Reported simulation times were zeroed at the start of the production run, thus excluding the 10 ns equilibration time.

7.3.2.2 Snapshot Selection

We selected snapshots from the MD simulation that met with which we predicted to be adequate criteria for reactivity, based on interatomic distances that most probably would lead to catalytically relevant barriers. The three criteria we used are: (1) the distance of the nucleophilic aspartate (closest oxygen) to the scissile carbon should be under 3.5 Å, (2) structural hydrogen bonds between D300 and the hydroxyl groups attached to the C2 and C3 carbon atoms in **Figure 7.1** should be under 2.5 Å (described previously as important to keep the sugar rings in the correct position for catalysis [235]) and (3) the existence of a water molecule with a proton within 2.5 Å from the glycosidic oxygen. After calculating the barriers, we concluded that one of our criteria (number 3) was irrelevant.
If we had not restrained our MD simulation, we would have had to add a fourth condition, which would have been the distance of the acidic proton in E233 to the glycosidic oxygen under a given value (*e.g.* 2.5 Å). This restraint increases the frequency of catalytic significant conformations, allowing to explore their properties more efficiently. It does not affect the values of the calculated barriers. The restraint increases the overall reaction rate that could eventually be calculated from the sampled structures, something that we are not pursuing in this study. Out of the 42 selected snapshots, only 18 were used in our analysis. The remaining 24 were excluded due to difficulties in characterizing the stationary points (Reactants or TS), or in guaranteeing that they lie in the same global minimum with respect to all degrees of freedom orthogonal to the reaction coordinate.

7.3.2.3 ONIOM

The structures obtained from the selected snapshots were studied using the ONIOM approach [205]. We used B3LYP [115,316] with the 6-31G(d) basis-set [125] for geometry optimization of the high layer and B3LYP, M06, M062X and MPW1B95, a higher basis set (6-311++G(2d,2p)) and correction for dispersion (D3) of its single point energy calculations. Atoms in the low layer were described by ff99SB [311]. The truncated bonds in the QM/MM boundary were capped with hydrogen link-atoms. The electrostatic interaction between the QM and MM layers was calculated within the electrostatic embedding formalism.

The high layer included two glucose monomers — before and after the scissile O-glycosidic bond, the neutral E233, the nucleophile D196 and structural D300, and the solvent water closest to the glycosidic oxygen. Several different water molecules occupied this position, so we used cpptraj [317] and custom scripts to find which water was the closest at each recorded snapshot. The 1000 closer waters to any protein atom were kept. Residues beyond 15 Å from the high layer were frozen, reducing the degrees of freedom during geometry optimizations. This procedure was found to be innocuous in recent studies [318]. All water molecules except the one closest to the glycosidic oxygen were frozen. Frozen water molecules within 3 Å from the high layer were removed to avoid artificial geometrical constraints, as we were not interested in studying the effect of local minima on the second solvation layer.

All relevant coordinates were freely optimized at the TS, without any bias. Nuclear vibrational frequencies were calculated to confirm the absence of imaginary frequencies in minima and one imaginary frequency at each transition state. Zero-point energies were computed at the B3LYP/6-31G(d) level of theory, using the harmonic oscillator/rigid rotor/particle in a box formalism [144,145]. Corrections for dispersion were calculated using Grimme's D3-DFT method [123] at each stationary point. Single-point energy calculations were performed using different density functionals (B3LYP, M06, M062X and MPW1B95) and a higher basis set (6-311++G(2d,2p)). The choice was based in a previous benchmarking of density functionals for the kinetics and thermodynamics of the glycosidic bond hydrolysis by glycosidases [319]. Final results are represented as electronic energies with zero-point energies and dispersion corrections at the M06-2X/311++G(2d,2p)-D3:ff99SB level of theory.

In the supporting information we provide activation energies including rigid/rotor harmonic oscillator entropies and thermal corrections. These were excluded from the main text because the contribution of rotational/vibrational entropy to the differences between barriers coming from different enzyme conformations (up to 3kcal/mol) is much smaller than the differences in internal energy among the barrier themselves. Therefore, an energy analysis of multiple individual conformation id appropriate for this study, avoiding technical problems associated with the calculation of entropy in constrained systems. The Gaussian 09 software was used to carry out all calculations [135].

7.4 Results and Discussion

7.4.1 Energies and kinetics

We calculated the activation energy of the glycosylation step in HPA based in 18 different conformations of the HPA-G5 complex sampled by during the MD simulation. We obtained activation energies ranging from 9.8 kcal/mol to 28.6 kcal/mol calculated at the M06-2X/6-311++G(2d,2p)-D3:ff99SB level of theory. Other functionals provided a similar picture with very similar activation energy spans (the values calculated with other density functionals are given in SI). Activation energies are reported in **Figure 7.2** and in **Table 7.1.** Other studies on enzymatic catalysis have found also a wide range of activation barriers [71,291,301-303,320-322].

Importantly, we observed that the barriers changed many orders of magnitude faster than the turnover rate. In fact, they changed at the ns timescale. For example, at 68.7 ns the barrier was 9.3 kcal/mol and 1.3 ns later, at the total 70.0 ns, the barrier increased to 20.5 kcal/mol. At the nanosecond timescale, where barrier fluctuations occur, conformational changes are mostly associated with thermal vibrations of bonds and angles, and rotamer transitions, and small movements of water molecules. There are no significant folding changes at this timescale. So, the enzyme responds to minimal thermal vibrations with enormous fluctuations in the reaction rate, due to a large number of interacting atoms. Since the turnover rate is slower than the barrier fluctuations, observed kinetics are a consequence of a few low activation barriers — occurring at specific conformations with critical interactions for catalysis — and the frequencies at which such low barrier conformations are visited, i.e. the partition of reactive conformers which can, in principle, be estimated through the accomplishment of specific geometric definitions that define the reactive enzyme conformations.

7.4.2 Structural analysis

To understand the structural reasons underlying the fluctuations in the activation barrier, we superimposed the structures of reactants (**Figure 7.3**) and transition states (**Figure 7.4).** The first half of the reaction cycle followed a well-known mechanism, with the glycosidic oxygen being protonated by E233, leading to the breaking of the glycosidic bond and formation of a transient carbocation at C1, which suffered nucleophilic addiction by D196 in a concerted dissociative step.

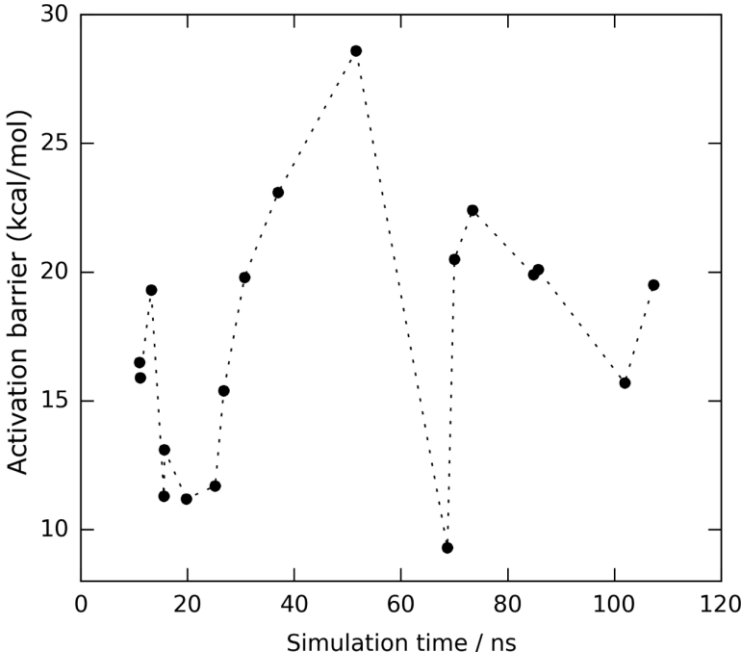


Figure 7.2 Activation barrier for snapshots selected from the MD simulation (the values are represented as zero-point corrected total energy $E_0^{\ddagger}$, calculated at M06-2X/6-311++G(2d,2p)-D3:ff99SB). The lowest activation barrier was found at 68.7 ns, and was 9.3 kcal/mol. The largest barrier of 28.9 kcal/mol corresponded to the snapshot recorded at 51.6 ns. There is a subtle tendency for structures closer in time to display similar activation barriers, but large variations occurred at a nanosecond timescale. The dashed line does not represent an interpolation of the values of the barriers, providing only visual guidance into the chronological order of snapshots.

Table 7.1 Activation barriers (zero-point corrected total energy $E_0^{\ddagger}$, calculated at the M06-2X/6-311++g(2d,2p)-D3:ff99SB) and relevant distances associated to the formation of reactive enzyme conformations.

time/ns	$\Delta E_0^{\ddagger}$ / kcal·mol^{-1}	d^R_{wat} /Å	d^R_{acid} /Å	d^R_{nuc} /Å	d^{TS}_{wat} /Å
11.0	16.5	3.68	2.95	3.44	2.36
11.2	15.9	3.03	2.95	3.43	2.11
13.2	19.3	3.75	3.00	3.30	2.58
15.6	11.3	2.30	1.94	3.21	1.98
15.7	13.1	2.64	1.95	3.17	2.14
19.8	11.2	2.36	2.76	3.29	2.01
25.2	11.7	2.91	2.74	3.14	2.09
26.8	15.4	2.90	2.54	3.57	2.08
30.7	19.8	2.52	2.67	3.70	2.07
37.0	23.1	4.40	2.83	4.10	2.17
51.6	28.6	3.29	2.79	3.44	2.10
68.7	9.30	2.92	1.89	3.24	2.08
70.0	20.5	4.42	2.60	3.85	4.34
73.4	22.4	3.53	2.63	3.67	2.13
84.8	19.9	3.83	2.54	3.53	3.61
85.7	20.1	2.89	2.48	3.47	2.14
101.9	15.7	2.63	1.98	3.14	2.09
107.3	19.5	4.44	3.15	3.51	2.11

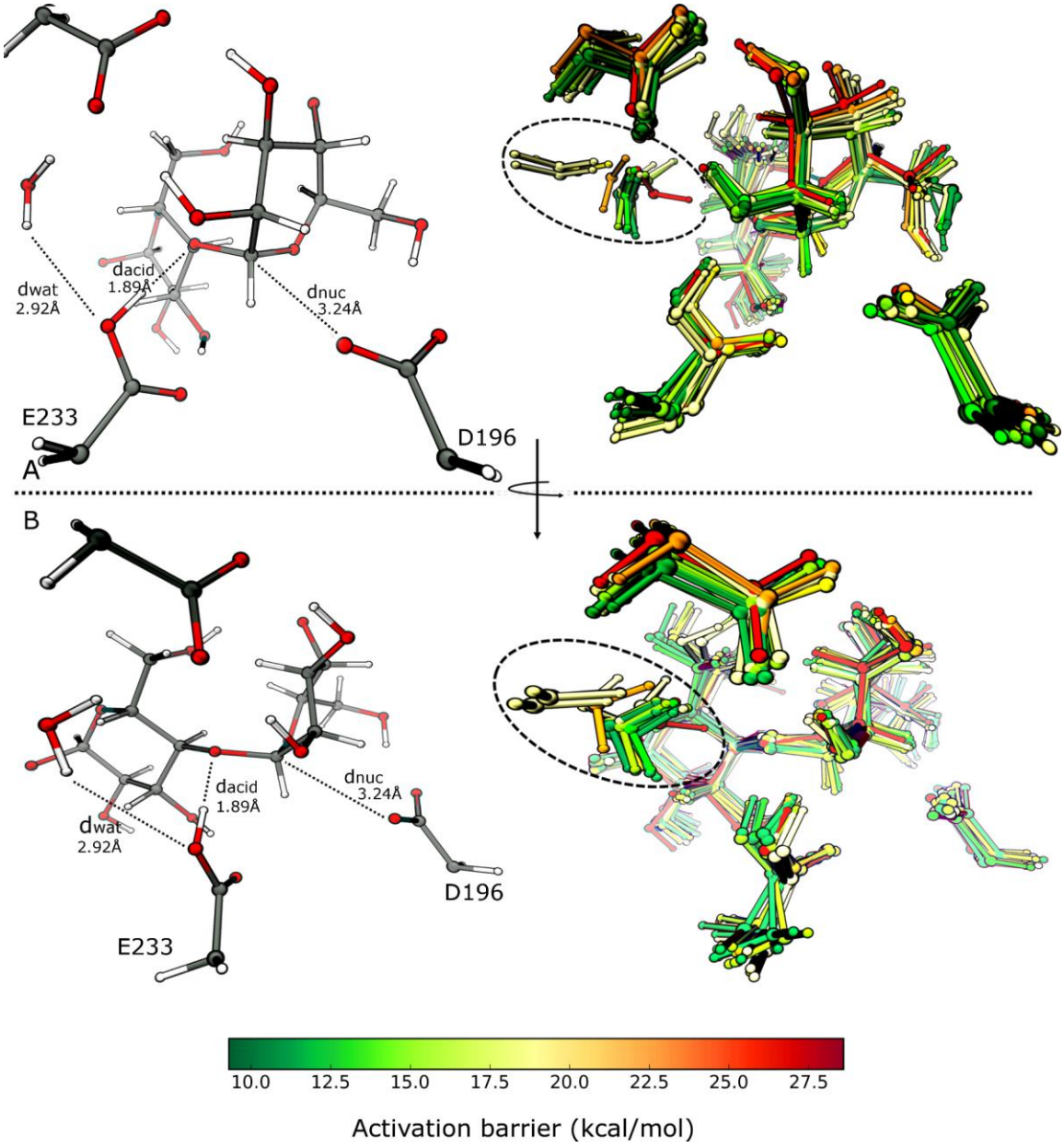


Figure 7.3 Reactant structures at the B3LYP/6-31g(d):ff99SB level of theory. Panels A and B represent the same structures rotated by about 60º. In each panel, the structure from 68.7ns, which is associated with the lowest energy, is represented along with the superimposed structures (for visual guidance). The water molecule can adopt a variety of positions and interactions as is highlighted by the dashed ellipse. Important distances d_{wat}, d_{acid} and d_{nuc} are represented with dashes. The structures are colored according to their associated activation barrier calculated at the M06-2X/6-311++g(2d,2p)-D3:ff99SB level of theory. A relationship between the position of the water molecule at the reactants and the activation barrier is evident. Overall, reactant structures do not align as well as transition state structures (see Table 4).

It is important to note that each React/TS pair lies on the same global minimum as could be verified by the superposition of the structures and the energy profiles along the reaction coordinate. Transition state structures displayed better superimposition than reactant structures, indicating that the conformational landscape of transition states around the active site is narrower than that of the reagents. This observation suggests a negative

conformational activation entropy ($\Delta S^{\ddagger}$) for the glycosylation step in α-amylase. It is interesting to note that the fact that TS structures are so confined around a well-defined conformation may be of great utility for the design of transition state analogues.

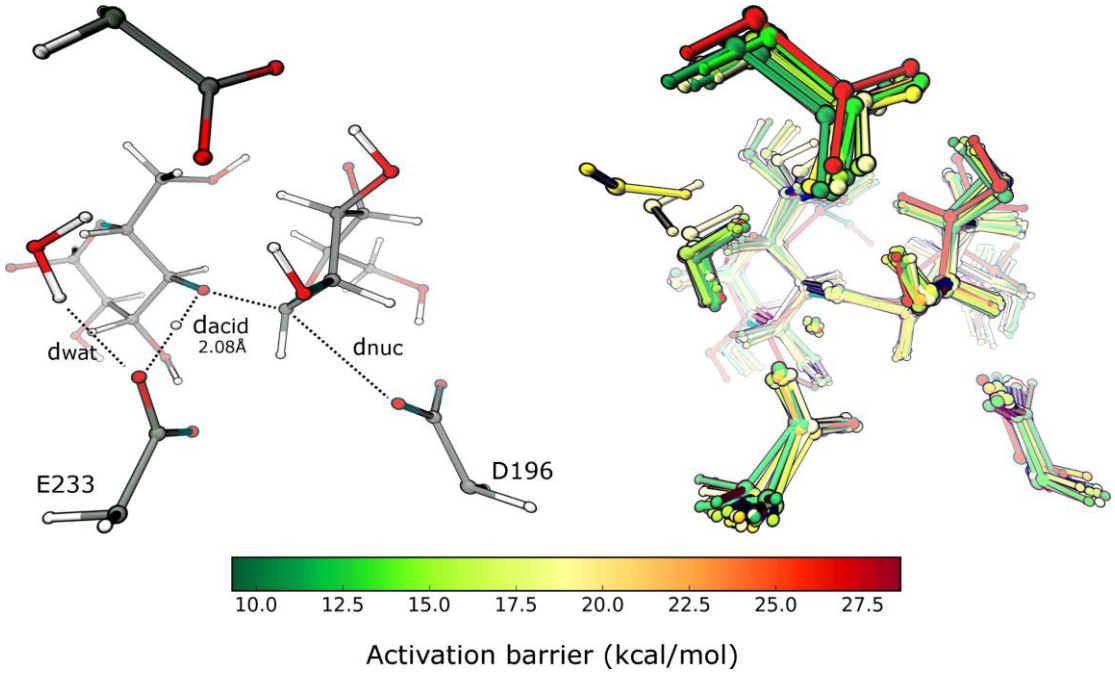


Figure 7.4 Transition state structures at the B3LYP/6-31g(d):ff99SB level of theory. The structure from 68.7 ns, which is associated with the lowest energy, is represented along with the superimposed structures for visual guidance. Important distances d_{wat}, d_{acid} and d_{nuc} are represented with dashes. The structures are colored according to their activation barrier calculated at the M06-2X/6-311++g(2d,2p)-D3:ff99SB level of theory. Transition state structures display better alignment than reactant structures (see Table 3).

In **Table 7.1,** the dependence of $\Delta E_0^{\ddagger}$ on the most relevant interaction distances is displayed (superscripts R and TS indicate if the distances were calculated at reactants or transition states, respectively). Low activation barriers only occurred if a set of geometric conditions were verified: short distances for d_{wat}, d_{acid} and d_{nuc} (see **Figure 7.1** and **Figure 7.5**).

There is a trend between the activation barrier heights and each of the three individual distances. This correlation becomes very evident if we take the three distances together, through a collective variable $d_{actsite}=d_{acid}+d_{nuc}+d_{wat}$ (**Figure 7.5** – panel E). This variable mirrors the closeness of the reactive residues to the substrate. An "expanded" active site will have larger d_{acid}, d_{nuc} and d_{wat} distances, whereas a "contracted" active site will have shorter d_{acid}, d_{nuc} and d_{wat} distances. It is clear that the degree of contraction of the active site (small $d_{actsite}$) controls much of the observed barrier height, and consequently enzymatic activity. This is also consistent with the results of an energy decomposition of this system (**Table 7.4** in SI), which shows that, in general, most of the fluctuation between different barriers derives from the different electronic energy barriers of the QM subsystems.

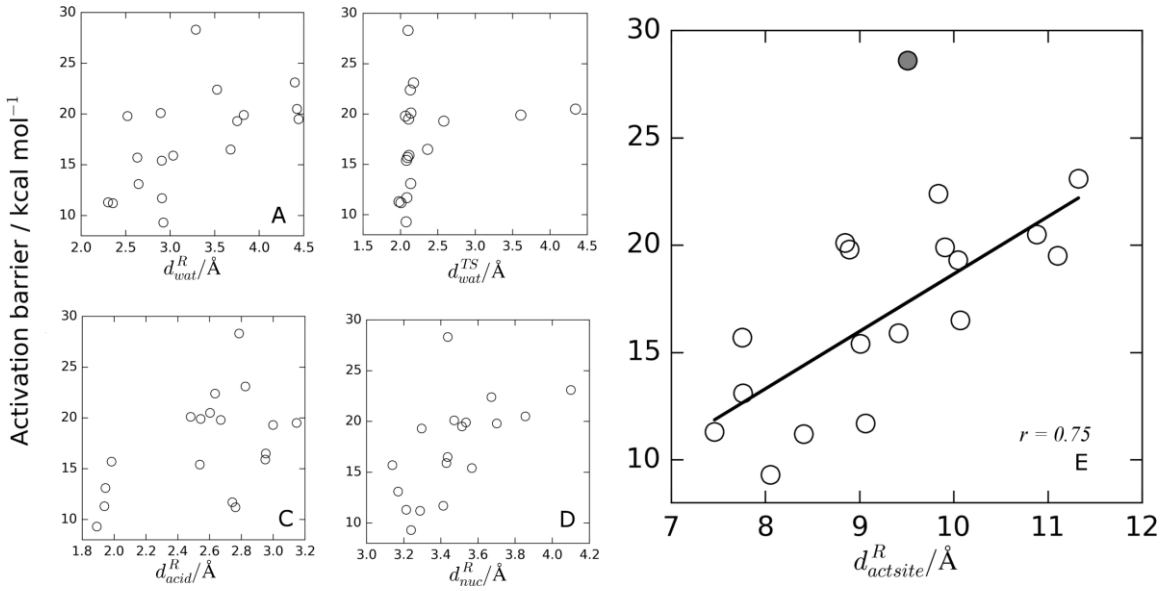


Figure 7.5 Correlation between a set of selected distances (d_{wat}, d_{acid}, d_{nuc}) and the corresponding activation barriers, calculated at the M06-2X/6-311++G(2d,2p)-D3:ff99SB level of theory. Panels A and B correspond to activation barrier for different d_{wat} at the R and TS, respectively. Panels C and D correspond to the correlation between d_{acid} and d_{nuc}, at the R, and corresponding activation barriers. A small trend between these catalytically-relevant distances at the reactants (Panels A, C and D) and the activation energy can be seen. In general, the small distances associated with small energies, and large distances associated with large energies. The same trend is not found at the TS (panel B), where d_{wat} span a much narrower range of values. The correlation becomes evident for the collective variable $d_{actsite}=d_{acid}+d_{nuc}+d_{wat}$ (panel E), that corresponds to a sum of the variables of panel A, C and D. This variable mirrors the closeness of the reactive residues around the substrate.

Two of these distances were expected to be important as they are implicated in the reaction coordinate — d_{nuc} corresponds to the nucleophile attack of D196 to the carbon in the scissile glycosidic bond (C1 in **Figure 7.1**) and d_{acid} corresponds to the proton transfer from E233 to the glycosidic oxygen. The importance of the third distance (d_{wat}) was surprising because it is not directly involved in the reaction but corresponds instead to a hydrogen bond established between a buried water molecule and residue E233. This hydrogen bond stabilizes the TS because the E233 carboxylate becomes negatively charged after donating its acidic proton to the glycosidic oxygen. In other words, the hydrogen bond lowers the pKa of E233, and helps E233 to fulfil its role as a general acid. $\Delta E_0^{\ddagger}$ depends heavily on d_{wat}, as all TS structures with low barriers displayed this required hydrogen bond with d_{wat} < 3.5 Å (see panel B in **Figure 7.5**).

Surprisingly, the distance of this hydrogen bond was as important for the activation barrier as the distances involving the reacting atoms. The correlation between d_{wat} and $\Delta E_0^{\ddagger}$ was so evident that this interaction must be sustained most of the time for HPA to be efficient and should be included in reactive conformations for this enzyme.

The strong dependence of the activation barriers for the whole enzymatic system on three specific coordinates (d_{wat}, d_{nuc}, and d_{acid}) revealed that, above anything else, these

distances must be below a certain threshold for the reaction to occur, otherwise the activation barrier is too large. Thus, we can significantly improve our definition of what constitutes a reactive conformation and estimate the frequency of reactive conformations during the MD simulation with improved accuracy. Since the d_{acid} distance was restrained in our simulation, all sampled microstates will display this interaction. In **Figure 7.6** we represent each recorded microstate of the MD simulation (small circles) as a function of d_{wat} and d_{nuc}. There is a significantly high number of frames displaying short distances for both interactions. The two interactions appear to behave independently from each other. The variation of these distances over MD simulation time is represented in Supporting Information. The values of these distances for the reactant state of ONIOM calculations are also represented (the color indicates the height of the barrier). It seems that HPA is efficient at sustaining these critical interactions.

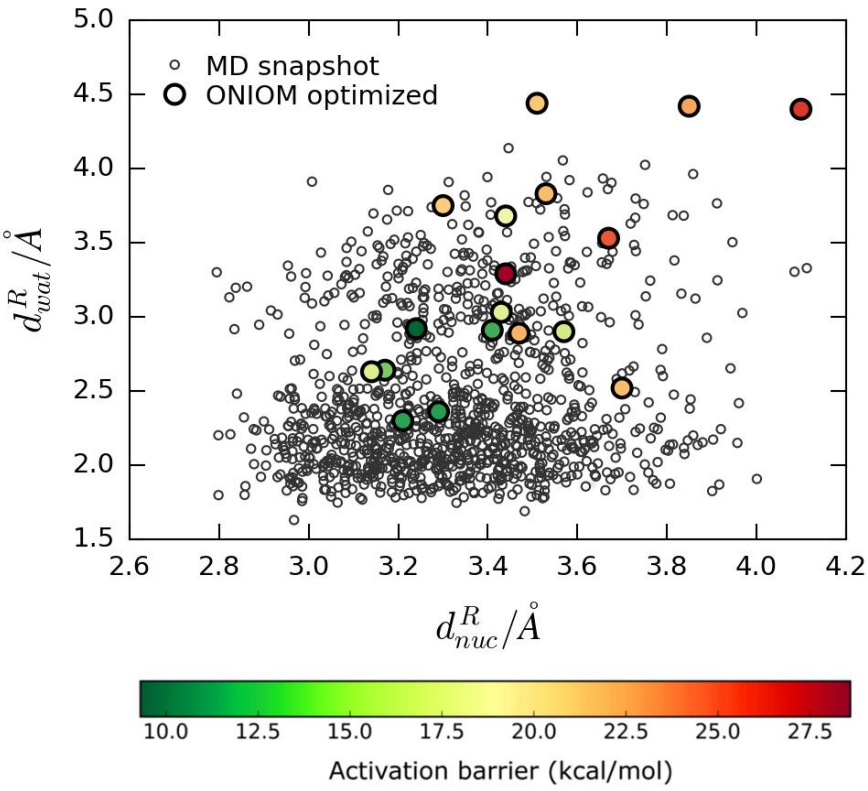


Figure 7.6 Distribution of d_{wat} and d_{nuc} distances during the 109 ns MD simulation of the solvated enzyme-substrate complex in the NPT ensemble. The same distances obtained after the QM/MM geometry optimization are also given and colored according to the obtained activation barrier. The proficiency of the enzyme in keeping these reactive enzyme conformations in place with a relevant frequency is evident.

Retrospectively, the third criteria we used for selecting snapshots from the MD simulation (the existence of a water molecule with a proton within 2.5 Å from the glycosidic oxygen) significantly reduced the number of selected snapshots that could display low activation

barriers. The water molecule spends much more time interacting with the neutral E233 than with the glycosidic oxygen. In the MD simulation, this interaction is observed in 60% of the microstates (d_{wat} < 2.5 Å), being small than 2.0 Å in 20% of them. This directly translates in a much higher number of reactive microstates, well beyond the 42 snapshots chosen due to the anticipated criteria. Importantly, the crystallographic structure '1CPU' displays this interaction at 1.9 Å distance (assuming O-H bond length equal to 1 Å), supporting the results we found in this study about the importance of this hydrogen bond for catalysis. Consistently, the ONIOM activation energy for the X-ray structure, calculated in a previous work [235], was close to the lower limit of the sampling performed here. The barrier of 10.0 kcal/mol, calculated at the B3LYP/6-311++G(2d,2p):ff99SB level of theory with dispersion corrections, was similar to our lower value calculated at the same theoretical level (see SI-**Table 7.3**). The main distances, d_{acid} and d_{nuc}, were also in the range that we obtained in the present work (d_{acid} = 2.70 Å and d_{nuc}=3.53 Å).

The x-ray structure represents the most common conformation of the crystallized system. The extent to which it represents a very relevant conformation, or the more common conformation, for the physiologic system depends mostly (but not only) on the reorganization induced by the bound ligands in the crystal, due to their dissimilarities in relation the natural substrates. Despite these limitations, whose important depends on the specific system under study, it has been widely observed that the activation barriers derived from this single conformation, more often than not, reproduce remarkably the observed activation energies (or the results of significant sampling), and here the situation was no different.

7.4.3 Conclusions

We identified important interactions in HPA that are associated with catalytic efficacy. Conformations of the HPA-G5 complex that display these important interactions at short distance have low activation barriers as calculated at the M06-2X/6-311++G(2d,2p)-D3:ff99SB level of theory. This data may be useful to devise accurate definitions of reactive enzyme conformations, which in turn can be used to estimate reactivity ($\Delta E^{\ddagger}$) from classical MD simulations alone, provided that a reasonable value of ΔG_{Chem} (the energy from a reactive reactant state to the TS) has already been computed by a suitable level of theory.

The importance of a hydrogen bond between a buried water molecule and amino acid residue E233 was surprising. This hydrogen bond lowers activation barriers by stabilizing the proton transfer from E233 to R as observed at the glycosidic oxygen in the scissile

bond. The essential pKa of the general acid E233 is basically controlled by this volatile hydrogen bond. The contribution of this interaction to the activation barrier is so significant that the water molecule ends up acting as a switch, tuning the reactivity of the enzyme the pace of its ultrafast movements. The volatility of this interaction (and also d_{nuc} in our MD trajectory) suggests that activation barriers have wide fluctuations on the nanosecond timescale, or even faster.

7.5 Supporting Information

The supporting information of this manuscript is presented below, and it has some additional results to complement the work presented above.

Table 7.2 (SI) Activation free energies calculated at the M06-2X/6-311++G(2d,2p):ff99SB level of theory, with D3 dispersion correction. Imaginary frequencies and contributions from zero point energy (ZPE), thermal and entropic corrections to the ONIOM activation barriers, calculated at the B3LYP/6-31G(d):ff99SB level of theory. Dispersion correction (D3) used for the M06-2X functional. All values are represented in kcal/mol.

			Corrections	
Time / ns	**ΔG‡**	**Imaginary frequencies**	**ZPE + Thermal energy + TΔS**	**Grimme Dispersion D3**
11.0	16.4	-180.2	-2.9	-0.3
11.2	18.2	-131.8	0.7	-0.1
13.2	19.4	-141.7	-2.5	-0.3
15.6	12.7	-153.7	-0.7	0.0
15.7	15.0	-108.0	0.6	-0.2
19.8	13.5	-201.0	0.8	-0.1
25.2	12.9	-156.5	-0.2	-0.3
26.8	16.2	-157.1	-0.9	-0.1
30.7	19.8	-160.1	-2.1	-0.5
37.0	22.4	-101.2	-4.6	-0.5
51.6	29.3	-153.5	-1.4	-0.2
68.7	11.5	-153.0	0.3	0.0
70.0	20.4	-158.4	-3.1	-0.1
73.4	21.9	-207.2	-2.9	0.1
84.8	20.6	-196.9	-2.6	0.1
85.7	23.1	-209.2	1.9	0.0
101.9	17.3	-96.1	0.1	0.0
107.3	20.1	-207.7	-2.2	-0.3

Table 7.3(SI) Activation energies obtained with four density functionals and the 6-311++G(2d,2p) basis set. $\Delta E_0^{\ddagger}$, corresponds to the activation internal energy with zero-point energy and corrections for dispersion (D3). $\Delta G^{\ddagger}$ corresponds to the activation free energy with zero-point energy, thermal and entropic corrections and dispersion correction (D3). All values are represented in kcal/mol. The low barriers were obtained with B3LYP, however it is known that this functional underestimated the barriers by ~3 kcal/mol in this type of reactions [319]. Having this into account we discussed the barrier obtained with the M06-2X functional.

Time / ns	**B3LYP**		**M06**		**M06-2X**		**MPW1B95**	
	$\Delta E_0^{\ddagger}$	$\Delta G^{\ddagger}$	$\Delta E_0^{\ddagger}$	$\Delta G^{\ddagger}$	$\Delta E_0^{\ddagger}$	$\Delta G^{\ddagger}$	$\Delta E_0^{\ddagger}$	$\Delta G^{\ddagger}$
11.0	15.9	13.7	18.3	18.0	16.5	16.4	17.8	16.6
11.2	16.0	13.9	15.9	17.7	15.9	18.2	18.0	18.0
13.2	16.9	14.9	19.9	19.8	19.3	19.4	19.7	18.6
15.6	11.2	10.4	13.1	14.5	11.3	12.7	12.9	12.9
15.7	9.9	10.5	13.0	14.8	13.1	15.0	13.3	14.5
19.8	10.9	10.0	11.4	13.4	11.2	13.5	12.8	13.5
25.2	13.4	11.1	12.9	14.1	11.7	12.9	14.7	13.9

Time / ns	B3LYP		M06		M06-2X		MPW1B95	
	$\Delta E_0^{\ddagger}$	$\Delta G^{\ddagger}$	$\Delta E_0^{\ddagger}$	$\Delta G^{\ddagger}$	$\Delta E_0^{\ddagger}$	$\Delta G^{\ddagger}$	$\Delta E_0^{\ddagger}$	$\Delta G^{\ddagger}$
26.8	14.1	14.0	15.5	16.2	15.4	16.2	17.2	17.6
30.7	18.9	18.4	20.5	21.0	19.8	19.8	21.6	21.8
37.0	18.9	19.7	18.8	17.5	23.1	22.4	23.5	22.8
51.6	28.7	28.9	31.3	31.8	28.6	29.3	30.1	30.7
68.7	9.9	10.0	11.0	13.1	9.3	11.5	11.4	12.3
70.0	20.2	16.2	21.1	19.1	20.5	18.9	22.6	19.9
73.4	23.8	21.4	21.5	20.9	22.4	21.9	26.0	24.5
84.8	17.8	16.0	20.3	21.9	19.9	20.6	21.5	20.9
85.7	18.1	18.8	20.0	22.7	20.1	23.1	22.2	24.1
101.9	12.6	10.9	14.1	15.3	15.7	17.3	15.5	15.3
107.3	22.9	27.8	19.9	21.5	19.5	20.1	24.2	22.9

Table 7.4 (SI) Activation energies of the QM subsystems in vacuum (QM barrier), compared with ONIOM activation energies ($\Delta E^{\ddagger}$), without ZPE correction. The values were calculated with the M06-2X/6-311++G(2d,2p):ff99SB functional and they are represented in kcal/mol. In general, most of the fluctuations between barriers comes from the energies of the QM subsystems, giving support to our conclusions that the contraction of the active site (small distances in the active site) controls the barrier of the reaction. The main exception was verified for the barrier observed at 73.4 ns. The low barrier observed in the QM partition is justified by the correct orientation of the water molecule to the aspartate. In this case the biggest contribution to the final energy comes from the atoms outside the QM layer.

Time/ns	QM Barrier	$\Delta E^{\ddagger}$
11.0	21.0	19.7
11.2	13.9	17.6
13.2	21.0	22.2
15.6	6.4	13.4
15.7	10.5	14.6
19.8	12.8	12.8
25.2	11.9	13.4
26.8	16.5	16.9
30.7	18.7	22.4
37.0	16.6	27.5
51.6	26.8	30.9
68.7	5.8	11.3
70.0	19.3	23.6
73.4	7.2	24.7
84.8	18.6	23.1
85.7	13.2	21.3
101.9	11.6	17.2
107.3	13.0	22.6

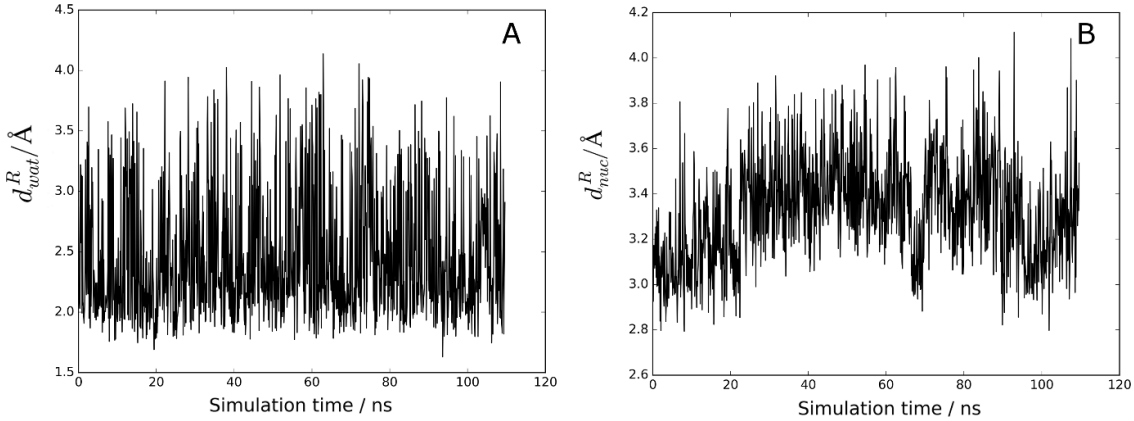


Figure 7.7 (SI) Variation of d_{wat} (A) and d_{nuc} (B) over MD simulation time.

CHAPTER 8. The influence of enzyme conformation on the HIV-1 protease catalytic mechanism

Ana Rita Calixto, Maria João Ramos and Pedro Alexandrino Fernandes
UCIBIO@REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências Universidade do Porto, Rua do Campo Alegre s/n,4169-007 Porto, Portugal

This work is being developed to explore further a previous subject studied by our group. In the previous work, it was showed that the rate of the reaction catalyzed by HIV-1 Protease is sensible to enzymatic structural fluctuations [71]. Here, we improve some methodological issues, that casted doubts in the previous results. Up to now, the results are similar to the previous published ones, showing that enzymatic fluctuations seem to have influence in the catalytic power of this enzyme, even though this new study uncovers new, interesting mechanistic hypotheses that were not fully understood in the previous work.

All the calculations presented here were performed by Ana Rita Calixto, as for the writing of the following draft of the manuscript. This work is still ongoing and some of the conclusion could change with the analysis of new results

8.1 Abstract

The role of conformational fluctuations on enzyme catalysis has been a matter of debate on recent studies devoted to the understanding of the origin of the catalytic power of enzymes. It is becoming clear that specific, instantaneous, not rare enzyme conformations, hiving the active site perfectly pre-organized for the reaction, led to low activation barriers. The present work is focused on exploring this aspect on the catalytic mechanism of HIV-1 protease (PR) with an adiabatic mapping method, starting from different initial structures, collected from a MD simulation.

The first step of HIV-1 PR catalytic mechanism was studied with the ONIOM quantum mechanics/molecular mechanics (QM/MM) methodology (B3LYP/6-31G(d):ff99SB), with final energies calculated at the M06-2X/6-311++G(2d,2p):ff99SB level of theory, in 30 different enzyme:substrate structures.

The results showed that the enzyme conformation influences, not only the energetic profile, but also the chemical progress of the reaction, following two possible mechanisms to obtain two different intermediates. Both mechanisms lead to the formation of a gem-diol intermediate, however the participation of the active site residues is different in each case. The structural organization of each residue in the active site seems to be important to the progress of the reaction. Very small variations on its orientation leads to differences in the barriers. There is not a single structural factor that accounts for the observed differences. However, some specific distances, and, consequently, the electrostatic interactions on the active site, are clearly associated with the progress of the reaction by one mechanism or the other. Current methodologies do not allow to determine which of the two mechanisms (if any) is dominant in physiologic conditions These results are important to understand the role of enzyme fluctuations on catalysis and give tracks to understand the origin of the catalytic power of enzymes.

Keywords

Enzymatic catalysis, Catalytic power of enzymes, Reactive enzyme conformations, Enzyme flexibility, HIV-1 Protease

8.2 Introduction

Enzymes are large molecules that have many degrees of freedom. Their structure fluctuates over time, resulting in many interchanging conformations [36,65,323]., There are many enzyme conformations well populated from which a reaction can start. In the same way, these reactants may lead to different transition states and products, or different energetics for similar transition states and products. The experimental reaction rate is usually made over macroscopic amounts of proteins and during macroscopic times, reflecting a weighted average over all these possibilities. A few single-molecule kinetic measurements have been experimentally made, and they confirm this picture [324-326].
Enzymatic studies using computational methods may easily be performed on single structures and isolate this effect that experimentally is extremely hard to measure

8.2.1 The influence of enzyme conformations on reactivity

Chemical reactions in enzymes only occur when the active site residues are in a suitable pre-organization (position, orientation) to promote the reaction. Otherwise, the energetic barriers will be very high, and the reaction will not take place. These positions and orientations fluctuate significantly in the ps-ns timescale, and even at larger timescales due to more global enzyme movements [327]. This strong dependence of the reactivity on the enzyme on the conformational orientation have been demonstrated by different theoretical works, in which the activation barriers were calculated for several different initial conformations of the same enzyme [71,299,328-335].
For example, in ketosteroid isomerase the barriers change in about 20 kcal/mol due to a structural variation. In fluoroacetate dehalogenase, the authors considered twenty snapshots for each of three different systems, and they found that the barrier of each reaction varies up to 15 kcal/mol. They associate these energetic fluctuations with structural parameters of the active site [321]. In the reaction catalyzed by P450, variations up to 17 kcal/mol were also found [336]. In a recent experimental work with enoyl-thioester reductases the authors also found that subtle changes in the geometric structure of the active site, are responsible for a dramatic loss of the catalytic power of this type of enzymes [337].
Our previous study on α-amylase showed that the position of a buried water highly influences the barrier of this enzyme. Values from 9.3 to 28.6 were found and they are dependent on the position and orientation of this water molecule and the instantaneous distance between the substrate and the two reactive residues [328]. Combining all these barriers into a single, observed barrier, is a matter of intense research [338]. The main

difficulty lies in the necessary extensive sampling that is needed to bring a proper weight to each of the individual barriers. However, the objective of these studies is not to calculate an accurate value for the "observed barrier" but instead to shed light into the structural requirements for a reaction to take place. Other methods, such as QM/MM molecular dynamics, implicitly deal and average these instantaneous barrier fluctuations. However, in these cases, the Hamiltonian used to describe the catalytic region must be very simplified in order to calculate the energies of all structures generated by these methods [263,339-343]. As the effect is accounted for implicitly, it cannot use to bring understanding about the nature of the interactions that stabilize the transition state, which is the major purpose of this study and other studies of this kind.

In the present work, we studied the first step of the catalytic mechanism of HIV-1 PR **(Figure 8.1**) trying to identify correlations between enzyme:substrate structural fluctuations and reaction mechanism/reaction rate. One of the possible approaches is to sample several reaction paths with QM(DFT)/MM methods, using different initial structures of the enzyme:substrate complex. Here, ONIOM QM/MM calculations were performed on HIV-1 PR, starting from different initial structures taking from a MD simulation. This model was selected for two main reasons: a) the size of this enzyme is small, and its structure is relatively simple and well-known, composed by two identical amino acid chains, with 99 residues each one; b) their catalytic mechanism is well studied, from both experimental and theoretical studies on this enzyme, as well as, from other studies in similar aspartic proteases [64,71,302,344].

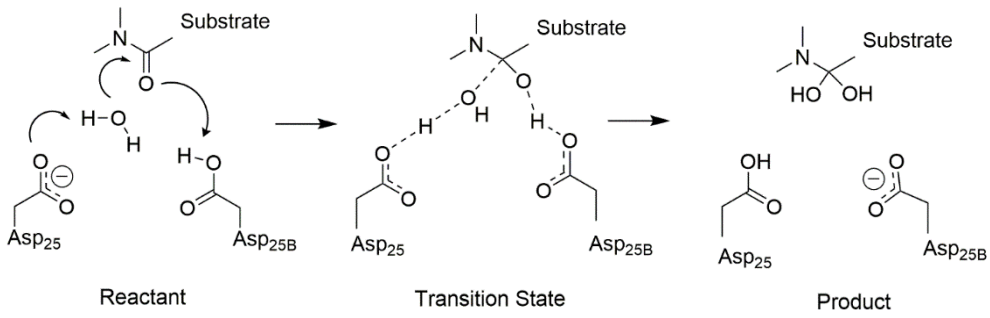


Figure 8.1 The first step of the catalytic mechanism of HIV1-PR, characterized by a nucleophilic attack of a water molecule on the carbonyl carbon of the substrate scissile bond, forming a tetrahedral intermediate.

The results were analyzing focusing on understand the relationship between activation free energies, obtained for each initial structure, and specific interactions that occur in each single conformation.

This enzyme was previously studies in our group, with the same focus [71], however, in the present work, some methodological aspects were modified: the number of atoms of the high layer (QM model) was increased to include the effect of the nearest groups treated

with a more accurate method; explicit solvent was included in the model; a very long equilibration time (100 ns) was used to allow for more extensive enzyme reorganozations. Here we used 30 snapshots of the complex enzyme:substrate from a MD simulation to study their reaction path, through QM/MM calculations. Then, we related the obtained barriers with structural parameters of each conformations, analyzing the main distances of the active site.

8.3 Methods

The computational protocol used here was very similar to the one used in one of our previous works [318] . The overall protocol consisted in the following steps: i) Modeling of the enzyme-substrate complex using the 4HVP PDB structure [279]; ii) Performing a small MD simulation to equilibrate the system; iii) Performing a large MD simulation to sample the conformational space of the system; iv) Studying the first step of the HIV-1 protease catalytic mechanism using QM/MM methods, in 30 different structures, collected from the MD simulation, and equally spaced in time (1 ns); v) Performing structural analysis of the active residues, correlating their structural fluctuations with the obtained activation energies.

The initial system was constructed starting from the 4HVP X-ray structure from the Protein Data Bank (PDB) [279] . This initial structure contains the complete enzyme complexed with a substrate-based peptide inhibitor with the sequence Ac-Thr-Ile-Nle-[CH2-NH]-Nle-Gln-Arg-amide. This structure was modeled to a correct substrate with the following sequence: Ac-Thr-Ile-Met-[CO-NH]-Met-Gln-Arg-amide, in the same way as in previous works [71,318]. A nucleophilic water molecule was added to the active center and the Asp_{25B} carboxylate was protonated. It is known that this residue needs to be protonated to initiate the catalytic mechanism of this enzyme.

Prior to the MD simulation, 9960 TIP3P water molecules were added to the protein:substrate complex [105], in a rectangular box of 88 Å x 67 Å x 71 Å. A minimum of 12 Å were left between any atom of the complex protein-substrate surface and the external molecules of the solvent box.

To equilibrate the modeled complex, we performed a first MD simulation without any restriction on the structure. In this simulation, the catalytic water molecule diffused away from the active site to the solvent and the catalytic aspartates turned their side chains to each other (the well-known very stable “resting state” of HIV-1 PR). To circumvent this, we forced the protein to adopt the less abundant catalytic conformation by constraining the distance between the catalytic hydrogen atom of $Asp25_B$ and the carbonyl oxygen

atom of the substrate. We used a harmonic potential having an equilibrium length of 1.80 Å between these two atoms and a force constant of 50 kcal.mol^{-1}Å^{-2}, using the same protocol as described in a previous work [71].

The PME method [284] was used to calculate the Coulombic interactions with the real part truncated at 10 Å. Explicit van der Waals interactions were also truncated at 10 Å. For the simulation without restriction we used the SHAKE algorithm [104] and a time step of 1 fs. In order to relax the system, removing possible tensions or clashes we started by a three-step minimization of the system using Amber 12 simulation package [285] with parm 99SB force field. First, the water molecules were minimized with the remainder of the system fixed. In these calculations the steepest algorithm for 5000 cycles and conjugated gradient algorithm for the last 5000 steps. Then the hydrogen atoms were minimized, fixing the remainder of the system (steepest descent algorithm to the first 5000 cycles, and conjugate gradient algorithm for the last 5000 steps). Finally, the position of all atoms was minimized (steepest descent algorithm to the first 15000 steps and conjugate gradient algorithm for the last 15000 steps). Starting from the structure obtained after the minimization procedure, we ran MD simulations, starting by an initial warm-up of the system from 0 to 300K during a 40 ps long simulation maintaining a constant volume and with periodic boundary conditions (canonical ensemble – NVT). Then we ran a MD equilibration on the whole system in the isothermal-isobaric ensemble (NPT) with the Langevin thermostat and isotropic position scaling, maintaining the temperature at 300 K and the pressure at 1 bar. Then the production dynamics was run during 200 ns with the same conditions.

After these steps, we selected different snapshots from the MD simulation based on time. We started by select some structures from the initial nanoseconds of the MD simulation, however, the results associated with these structures were discarded due to optimization problems or very high barriers. We associated these results to an incorrect position of the catalytic water molecule (not present in any x-ray structure), that was modelled in the active site by us. Taking this in account, we decided to select snapshots after 100 ns of MD simulation.

30 structures were selected and similar QM/MM models, applying an ONIOM scheme as implemented in the Gaussian 09 software package [135] were defined for each enzyme:substrate complex. The structures were sampled equally spaced in time (1ns). All the prepared systems contained a total of 6232 atoms, with 90 atoms in the QM layer and the remained system in the MM layer. The QM layer contained the two catalytic aspartates, the nucleophilic water molecule, seventeen atoms of the substrate, two structural water molecules and some residues around the groups that have an active participation on the reaction (Ala127, Gly126, Thr125, Gly27, Ala28, and the carbonyl

group of Thr_{26}). A cap of 1000 water molecules (~3Å around the protein) were kept in the model. The water molecules were frozen during all the calculations [318], except the catalytic and structural ones, which were present in the QM layer.

The interaction between the layers was treated with the electrostatic embedding scheme. The QM layer was optimized with the density functional B3LYP [115,206] and 6-31G(d) basis set. Hydrogen atoms were used as link atoms, where QM covalent bonds were truncated. The reaction path was studied in the same manner for all models, using the same reaction coordinate (the distance between the oxygen of the nucleophile water molecule and the carbonyl carbon of the scissile peptide bond of the substrate). The structures with the highest energy in the performed scans were used as starting guesses to optimize the geometry of the transition sate. Nuclear vibrational frequencies were determined to confirm the nature of the stationary points (absence of imaginary frequencies in minima and one imaginary frequency at each transition state). Zero-point energies were computed at the B3LYP/6-31G(d) level of theory [115,125,316], using the harmonic oscillator/rigid rotor formalism [144,145]. Intrinsic reaction coordinate calculations (IRC) were performed to obtain reactant and product structures in the same relative minimum. Single point energy calculations were performed using M06-2X density functional and a higher basis set (6-311++G(2d,2p)). Final results were represented as electronic energies with the correction to the zero-point energy. All the calculations were performed using ONIOM scheme [205] as implemented in Gaussian 09 software package [135].

8.5 Results and Discussion

The MD performed in this work generated a Boltzmann distribution of different microstates, in the reactant state. We studied the barriers of a significant number of structures and used them as an ensemble of initial structures to study our reaction (**Figure 8.2).** The sampling obtained by classic MD simulation has some known limitations: 1) only the reactants microstates are sampled during MD time, because the force field is not able to describe electronic properties of the system (bonds breaking and formation); 2) Some enzyme rearrangements, that occur is a larger time scale, comparable to the time scale sampled, are not accessible. More than that, from the MD simulation to the QM/MM studies, there are some methodological differences that are important to have in mind: 1) first of all, the molecular model is different. While in MD simulation the system is studied as periodic, with explicit solvent, in QM/MM calculations, a single protein:substrate system (in a small cap of constrained water molecules) is used; 2) the Hamiltonian, which is used to treat the system, is different in both methodologies. In MD simulations, classical

molecular mechanics (force field) is used, whereas in QM/MM calculations the active site is treated with QM methods. Despite these limitations, the methodology used here is adequate to help us to understand the influence of structural fluctuation in the activation barrier. However, these factors, together with the (still) limited sampling, makes difficult to assign a proper weight to each of the calculated barriers.

Figure 8.2 shows the distribution of energies of the reactants states in the MD simulation (grey bars). The purple line indicates how many structures from each MD energy range we took to perform our calculations.

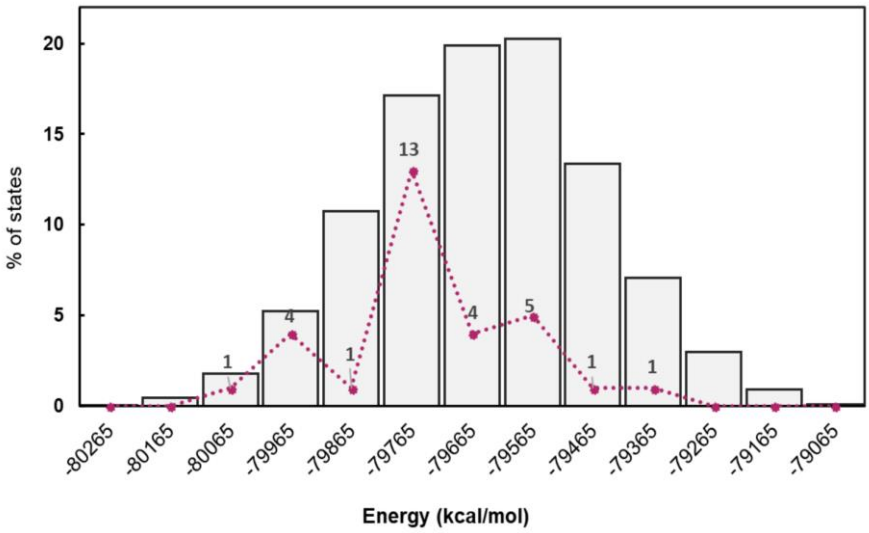


Figure 8.2. Energy distribution of the ensemble generated by the MD calculation (grey bars) and of the conformations taken for the QM/MM calculations (black line). The distribution peaks around the average energy and decays slowly in both directions, very well in line with the expected NPT ensemble distribution of energy, which results from the product of the Boltzmann factor, decaying very fast with increasing energy, and the density of states, increasing very fast with increasing energy.

Out of the 30 selected initial structures, only 19 were used into our analysis. The remaining 11 were excluded due to optimization problems and difficulties in characterizing the stationary points (reactants or transition states). Therefore, the activation energy associated with the nucleophilic attack of a catalytic water molecule on the carbonyl carbon of the substrate scissile bond were calculated based on 19 different structures, sampled during the MD simulation.

8.5.1 Dispersion of the activation barriers

Our results showed activation barriers ($\Delta E + ZPE$) ranging, from 17.3 to 32.2 kcal/mol at the M06-2X/6-311++G(2d,2p):ff99SB level of theory **Figure 8.3** and **Table 8.1**). Two

different mechanism were observed, starting from different structures. These mechanisms were represented by different colors in **Figure 8.3** and they were schematized in **Figure 8.4**. The results showed the barriers changed in the nanoseconds time scale. The energy barrier, obtained using the structure taken after 120 ns of MD simulation, was the lowest between all our measures, and corresponds to 17.3 kcal/mol. After 1 ns the activation energy increased to above 30 kcal/mol.

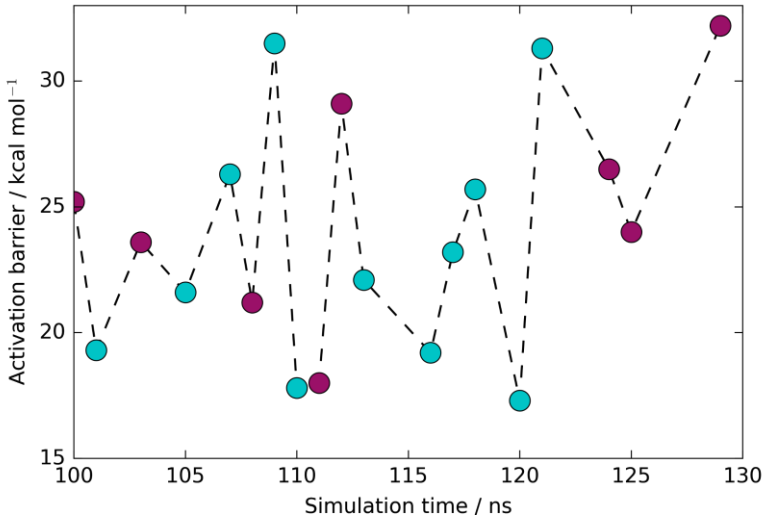


Figure 8.3 Activation barriers for 19 snapshots selected from the MD simulation. These barriers corresponded to zero-pointed corrected total energies ($\Delta E_0^{\ddagger}$), calculated at the M06-2X/6-311++G(2d,2p): ff99SB level of theory. The cyan and purple points correspond to structures that react through mechanism A or mechanism B, respectively. The lowest activation barrier (17.3 kcal/mol) was obtained, starting from the structure taking after 120 ns of MD simulation. Some variations occur in the nanoseconds timescale. The dashed line provides only a guidance for the chronological order of the barriers and does not correspond to an extrapolation of the energies between them.

Small changes in the orientation of the active site residues (bonds, angles and dihedrals) as well as, small movements of the water molecule, may justify the different barriers, consistently with the findings of previous works The conformational fluctuations do not correspond to changes in folding (which take place in much larger timescales), but instead to much subtle changes that, despite being small, can modify the very important chemical interactions between the substrates and the active site, leading to large changes in the reaction rate.

To understand if the observed fluctuations in the activation barriers were only related with the active site residues, we calculated the activation energies of the QM subsystem in vacuum (**Table 8.3** (SI)). As expectable, the large contribution to the ONIOM energy, came from the reactive atoms (QM layer). In general, most of the fluctuations came from the energies of the QM subsystem, given support to our previous observation

The turnover of HIV-1 PR (second timescale) takes place in a much slower timescale that the observed fluctuations of the barriers. It means that the experimental observed kinetics

could be a consequence of only a few, low activation barriers, that occur at well-defined conformations, with a specific active site highly tuned pre-organization. Despite the very limited sampling achieved here, due to the high-level theoretical methods employed, the frequency at which these low-barrier structures appears, (7 out of 19, with barriers smaller than 22 kcal/mol), that should not be very far from their ensemble probability, is more than enough to overcome the very high Boltzmann penalties associated to the more frequent, high-barrier structures. Thus, the former seems to determine the reaction's kinetics. The conclusion is consistent with other previous studies, where low-energy barriers were found not in most of the explored conformations but still in a very reasonable number of cases.

8.5.2 Different reaction mechanisms

The two different mechanisms observed in **Figure 8.4** were previously observed by our colleagues [71] . Mechanism A is well described and widely accepted in the literature for HIV-1 PR, as well as for other similar aspartic proteases [234,345]. It is characterized by a nucleophilic attack of a water molecule, present between both catalytic Asp residues, on the carbonyl carbon of the scissile bond, while it loses a proton to Asp25A. At the same time, the carboxylate of Asp25B protonates the carbonyl oxygen of the peptidic bond. In mechanism B, the Asp25A does not participate in the reaction. In this case, when the water molecule attacks the carbonyl carbon, the water proton is abstracted by the free oxygen of Asp25B, while the Asp25B acidic proton is transferred to the carbonyl oxygen. These two mechanisms have the "same chemistry", the difference is weather a single Asp residue acts as an acid and a base at the same step, or if the two functions are divided by two equivalent Asp residues.

Low barriers were found in both A and B mechanisms, provided that the conformation of the active site is adequate. For example, the structure taken after 109 ns of MD simulation is associated to a high barrier of 31.5 kcal/mol, which means that this structure is not properly to initiate the catalytic mechanism. However, only after 1 ns of MD simulation (110 ns), small fluctuations on the structure makes it able to perform the catalysis (activation barrier of 17.8 kcal/mol). For these two structures, the reaction the reaction takes place through mechanism A. The structure correspondent to the next ns (111 ns) reacted, in turn, throughout mechanism B with a plausible activation barrier of 18.0 kcal/mol. A structural analysis compared both mechanisms and correlate them with the activation barriers was performed and is detailed in the next topic.

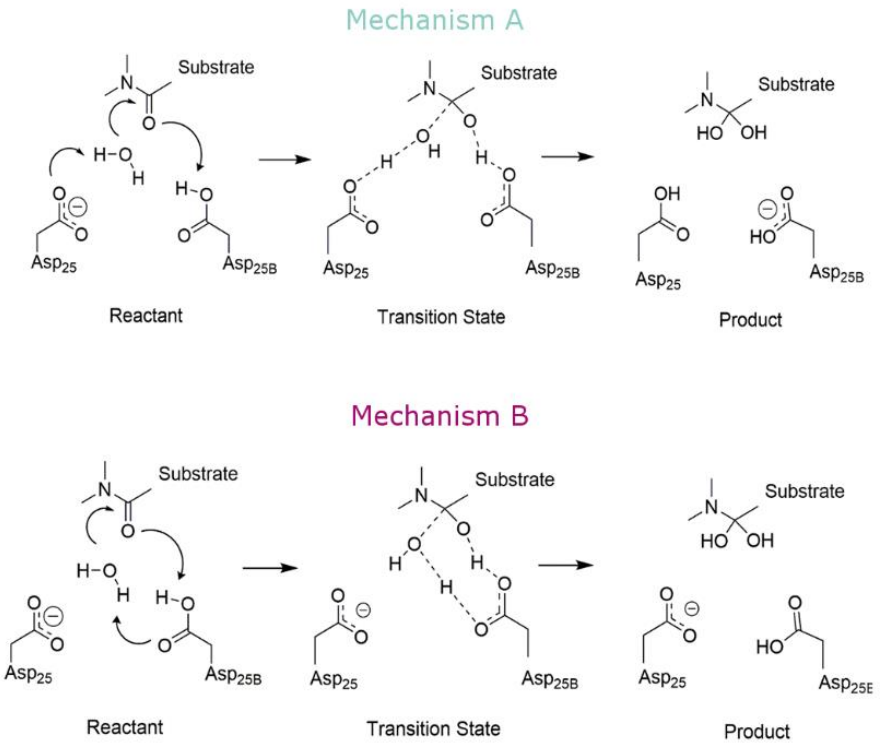


Figure 8.4 Different mechanisms observed for the first step of HIV-1 protease. Both mechanism A and B are characterized by a nucleophilic attack of the water molecule of the carbonyl carbon of the scissile bond, forming a tetrahedral intermediate. In mechanism A, which is well accepted, one of the two active site aspartates is ionized (Asp25B), while the other is protonated (Asp25), during the nucleophilic attack. In mechanism B the same aspartate residue (Asp25B) is ionized and protonated by the water molecule.

8.5.3 Structural analysis

Figure 8.5 represents the QM layer used in the QM/MM calculations, highlighting the main distances that were analyzed to understand the activation barrier fluctuations. As mentioned above, the QM and MM layers were constructed with the same number of

atoms, for all structures. For simplify, we only represent the QM layer from the structure that react through the smallest barrier (120 ns – 17.3 kcal/mol). Six distances were selected to analyze: $d1$ to $d4$ correspond to distances between reacting atoms, and $d5$ and $d6$ correspond to hydrogen bonds that tune the pKa of the reacting Asp25A residue. $d1$, which corresponds to the reaction coordinate, is the distance between the oxygen of the catalytic water and the carbonyl carbon of the substrate scissile bond (Met201); $d2$ is the distance between the hydrogen atom from the Asp25B carboxylic group and the oxygen atom from the carbonyl group of Met201; $d3$ corresponds to the smallest distance between an oxygen from the Asp25A carboxylic group and the catalytic water molecule; $d4$ corresponds to the distance between the non-protonated oxygen from the carboxylic group of Asp25B and the catalytic water molecule; $d5$, is the distance between Thr125 and an oxygen from the Asp25A and, finally, $d6$, is the distance between an oxygen atom from Asp25A carboxylic group and a water molecule that enters during the MD simulation.

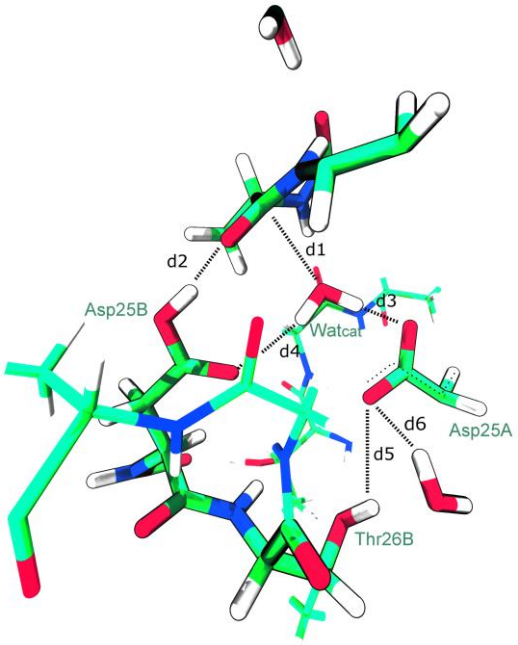


Figure 8.5 Reactant state from the structure taken after 120 ns of MD simulation, which is associated with the lowest energetic barrier. Only the QM layer was represented for simplicity. Important active site distances are highlighted: *d1*, established between the oxygen of the catalytic water molecule and the carbonyl carbon of the substrate scissile bond;*d2*, established between the hydrogen atom from the Asp25B carboxylic group and the oxygen atom from the carbonyl group of the scissile bond; *d3*, that corresponds to the smallest distance between an oxygen of the Asp25A carboxylic group and the catalytic water molecule; *d4*, established between the non-protonated oxygen from the carboxylic group of Asp25B and the catalytic water molecule; *d5*, established between Thr26B and an oxygen from the Asp25A, and *d6*, established between an oxygen from Asp25A and a non-catalytic water molecule. A dihedral angle formed between an oxygen atom from Asp25A carboxylic group, the following carbons (Cγ and Cβ) and the next hydrogen, was also evaluates (Dotted line).

Beyond these distances, the dihedral angle formed between one of the oxygen from the carboxylic group of Asp25A, the following carbons (Cγ and Cβ) and the next hydrogen, (which determines the rotation of the carboxylic group) was also evaluated, to evaluate the role of the Asp25S carboxylic group orientation on the catalytic mechanism of this enzyme. These structural parameters were evaluated in the optimized reactant structures (after IRC calculations) and in the optimized transition state structures.

Table 8.1 summarizes the results of this work, indicating the type of mechanism that corresponds to each selected snapshot and the obtained activation barriers (calculated at M06-2X(6-311++G(2d,2p):ff99SB level of theory. In **Table 8.2**, the evaluated distances were represented.

The obtained barriers raged from 17.3 kcal/mol to 32.2 kcal/mol. This range was similar for both mechanism: 17.3-31.5 kcal/mol for mechanism A and 18.0-32.2 kcal/mol for mechanism B. The barriers for the two mechanisms can be considered as equivalent within the accuracy of the method.

Table 8.1 Type of mechanism and activation barriers (Zero-point corrected Total Energy, $E_0^{\ddagger}$,calculated at the M06-2X/6-311++G(2d,2p)-D3:ff99SB level of theory) for each selected snapshot. The exponential average is also represented.

Time (ns)	Type of mechanism	$E_0^{\ddagger}$ (M06-2X(6-311++G(2d,2p)):ff99SB) kcal/mol
100	B	25.2
101	A	19.3
103	B	23.6
105	A	21.6
107	A	26.3
108	B	21.2
109	A	31.5
110	A	17.8
111	B	18.0
112	B	29.1
113	A	22.1
116	A	19.2
117	A	23.2
118	A	25.7
120	A	17.3
121	A	31.3
124	B	26.5
125	B	24.0
129	B	32.2
Exponential average (kcal/mol)		**18.7**

In the previous work on HIV-1 PR, our colleagues found the B alternative path only two times (between 39 structures) associating it with very high barriers. It is also interesting to note that, the fluctuations of the activation barriers, that we observed in this work, were on a smaller scale when compared to the previous ones [71]. These different results could be justified by the different preparation of the system. The inclusion of water molecules on the model, and the increase of the atoms treated with QM methods, can be enough to promote the observed differences.

Table 8.2 Type of mechanism and active site distances (Å) for each selected snapshot.

		Reactant						TS					
Time (ns)	**Type of mechanism**	**d1**	**d2**	**d3**	**d4**	**d5**	**d6**	**d1**	**d2**	**d3**	**d4**	**d5**	**d6**
100	B	3.05	1.63	1.87	3.05	1.70	2.09	1.76	1.02	2.67	1.29	1.72	2.26
101	A	2.97	1.69	1.63	2.97	1.79	4.16	1.86	1.44	1.07	2.01	2.02	4.04
103	B	2.95	1.64	1.80	2.95	1.82	1.78	1.68	1.02	1.87	1.39	1.93	1.85
105	A	2.79	1.68	1.71	2.79	4.48	1.87	1.68	1.01	1.49	1.62	5.12	1.89
107	A	2.98	1.66	1.78	2.98	3.05	1.94	1.68	1.05	1.38	1.63	3.41	2.78
108	B	2.74	1.68	1.73	2.74	1.81	1.68	1.70	1.02	1.91	1.32	1.90	1.71
109	A	2.92	1.66	1.80	2.92	1.77	1.99	1.74	1.02	2.11	1.28	1.81	2.05
110	A	2.78	1.69	1.78	2.78	3.47	1.83	1.69	1.04	1.42	1.62	3.80	2.13
111	B	2.90	1.68	1.79	2.90	1.82	1.72	1.67	1.02	1.65	1.44	1.95	1.78
112	B	2.75	1.68	1.70	2.75	2.75	2.05	1.92	1.07	1.22	2.05	2.82	3.25
113	A	2.89	1.67	1.69	2.89	1.77	3.04	1.76	1.33	1.21	1.88	1.88	3.51
116	A	3.00	1.65	1.86	3.00	3.58	2.50	1.61	1.02	1.56	1.48	3.72	3.58
117	A	2.84	1.67	1.71	2.84	1.74	3.12	1.74	1.03	1.32	1.74	1.90	3.54
118	A	2.91	1.68	1.77	2.91	1.74	1.85	1.73	1.02	1.89	1.28	1.80	3.03
120	A	2.75	1.68	1.70	2.75	2.75	2.05	1.92	1.07	1.22	2.05	2.82	3.25
121	A	3.04	1.62	1.73	3.04	1.78	1.78	1.77	1.03	2.45	1.27	1.84	1.72
124	B	2.80	1.65	1.76	2.80	1.79	1.65	1.64	1.02	1.76	1.39	1.91	1.67
125	B	2.82	1.67	1.78	2.82	1.79	1.75	1.68	1.01	1.85	1.28	1.86	1.75
129	B	3.02	1.66	1.73	3.02	1.85	1.85	1.80	1.02	2.41	1.29	1.87	1.88

To understand the origin of the observed fluctuations in activation barriers, all the individual distances of each structure were represented against the respective barrier. The results showed that there is no evident correlation between the individual distances and the energetic barriers (see **Figure 8.11** and **Figure 8.12** in Supporting Information). When the rotation of the Asp25A was evaluated, (**Figure 8.6**) no trend was observable between it and the activation barriers. However, some differences, between mechanism A and B, were visible. The occurrence of mechanism **B** was associated with two dihedral values, while the occurrence of mechanism **A** seemed to be independent of this dihedral angle. The orientation of Asp25A is defined by this dihedral angle. When it adopted those values (associated with mechanism **B**), its carboxylic group was stabilized by Thr125

and/or a structural water molecule (see also **Figure 8.9**). Consequently, a decrease of the pKa of this residue was observed, favoring the occurrence of mechanism **B**.

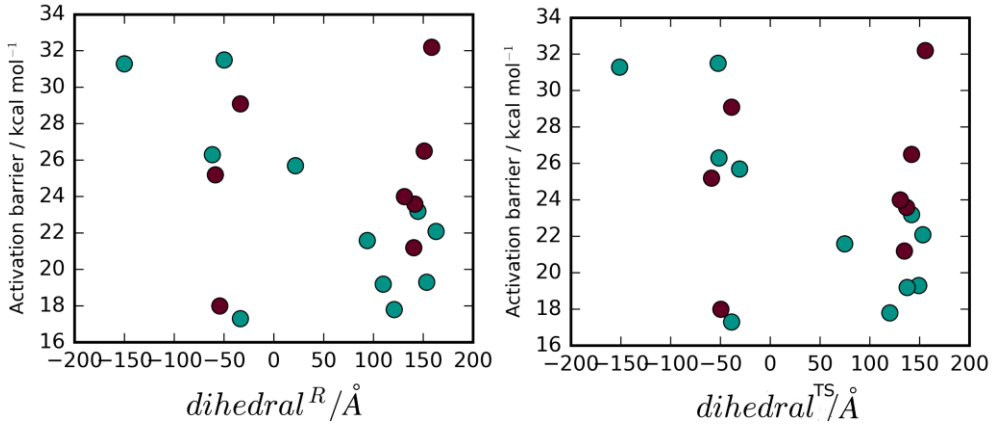


Figure 8.6 Correlation between the Asp25A dihedral angle and the corresponding activation barriers, for optimized Reactants and Transitions States. Blue and purple circles correspond to structures that react through mechanism A and B, respectively. The occurrence of mechanism B was dependent on the value of this dihedral (~-50 or ~150).

8.5.3.1 Influence of collective variables

Collective variable (sums of different distances) were evaluated to understand the reasons that lead the occurrence of each of the two mechanisms, and the values of the activation barriers.

Concerning the propensity for following mechanism A or B, the results showed that the strength of the hydrogen bonding to Asp25A is an important factor to define the progress of the reaction. Smaller values of the collective variable $d5 + d6$ mean stronger hydrogen bonding to Asp25A and consequently the lowering of the pKa of this residue, making more difficult its role as a base. In these cases, the Asp25B has more tendency to act as an acid and a base at the same time, given a proton to the oxygen carbonyl carbon and, receiving a proton from the water molecule (Mechanism B).

The progress of the reaction by this mechanism is also dependent on the strength of the hydrogen bonding of Asp25B to the water molecule (making easier its deprotonation by this residue) and the distance between the Asp25A and the water molecule, where a larger distance makes the deprotonation difficult. These two conditions are summarized in the collective variable d4-d3. Altogether, the collective variable $d5 + d6 + (d4 - d3)$ includes all the discussed aspects **(Figure 8.7).**

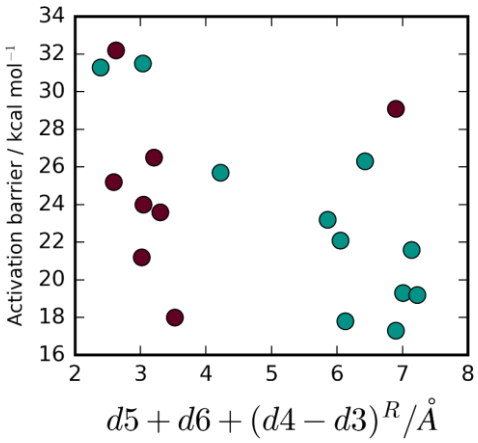


Figure 8.7 Correlation between the collective variable $d5 + d6 + (d4 - d3)$, the activation barriers and the propensity for following mechanism A or B. Blue and purple circles correspond to structures that react through mechanism A and B, respectively. It is visible a clear separation between the reactants that following mechanism A or B. The occurrence of mechanism B (purple) are related with small values of $d5 + d6$, that mean a stronger hydrogen bonding to Asp25A and, consequently, a decrease of its pKa. Small values of $d4$ and large values of $d3$ also controls the progress of the reaction by this mechanism. The opposite tendency is observed for mechanism A. When these conditions were not fulfilled the barriers were very high for both mechanisms.

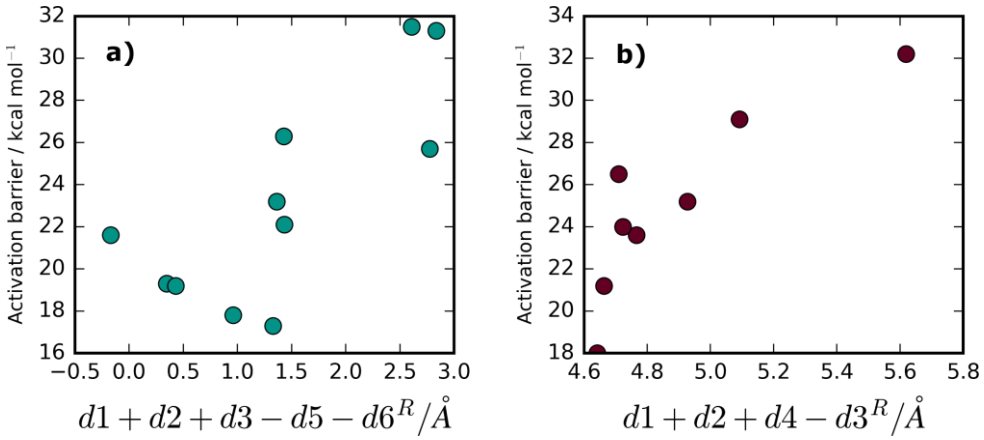


Figure 8.8 Correlation between the collective variable $d1 + d2 + d3 - d5 - d6$, for mechanism A, and $d1 + d2 + d4 - d3$, for mechanism B, and the activation barriers. Blue and purple circles correspond to structures that react through mechanism A and B.

Concerning the barrier fluctuations, different collective variables were selected for mechanism A and B and were represented (**Figure 8.8**). In both mechanisms A and B, the activation barrier was dependent on small values of the distances $d1$ and $d2$. However, for mechanism A, small values of *d3* and large values of d5 and d6 (weaker hydrogen bonds with Asp25A) are an important factor to promote the reaction. These results are summarized in the collective variable $d1 + d2 + d3 - d5 - d6$ (**Figure 8.8** – a)). High values of this collective variable lead to high barriers, making difficult the progress of the reaction. For mechanism B**,** beyond the influence of $d1$ and $d2$, the strength of the

hydrogen bonding between the catalytic water molecule and the Asp25B, is an important factor to define the barrier if this reaction, being small values associated with the smallest energies Large values of $d3$, along with the previous conditions also promoted the progress of mechanism **B** by small barriers. These results are summarized in **Figure 8.8**–b, in which the correlation between this collective variable $d1 + d2 + d4 - d3$ and the reaction barrier is evident.

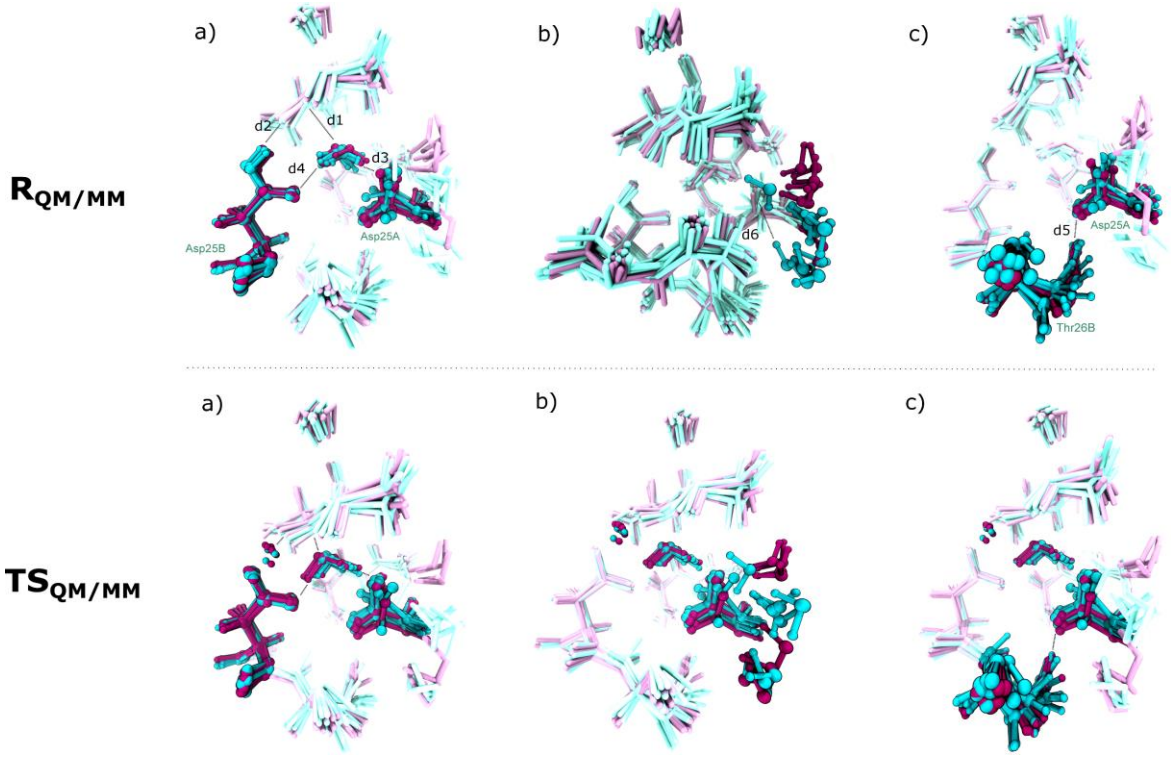


Figure 8.9 Superimposition of the structures after QM/MM optimizations, at the B3LYP/6-31g(d):ff99SB level of theory (Reactant and Transition State).The structures are colored according to the mechanism by which they react (Blue – Mechanism A; Purple – Mechanism B). The reactive residues (Asp25A, Asp25B and the catalytic water molecule) are highlighted in the first column (a); the structural fluctuations of the water molecule that is hydrogen bonded to Asp25A are highlighted in the second column (b); the structural fluctuation of the Thr26B sidechain is highlighted in the third column (c). Important active site distances were highlighted.

In **Figure 8.9** stationary structures (Reactant and Transition States) were superimposed and colored by mechanism (Mechanism A – Blue; Mechanism B – Purple). Three different regions of the active site were highlighted.

In the first column (a) the catalytic Aspartates and the catalytic water molecule were highlighted. Some differences in the Asp25A orientation were visible. Its orientation seemed to be more flexible in structures that react through mechanism A, as showed above in **Figure 8.6.**

The more evident difference between the structures that lead to Mechanism A or Mechanism B is the initial position of the water molecule, hydrogen bonded to Asp25A. The b) panels on of **Figure 8.9** show that in structures associated with mechanism B

(purple), the water molecule was closer to the Asp25A, while in blue structures (Mechanism A), the same water molecule was away from this residue. In c) panels, the differences on the orientation of the Thr26B sidechain was highlighted. As pointed above, when this residue is hydrogen bounding to Asp25A, the occurrence of mechanism B is favored.

In **Figure 8.10**, the structures were divided by mechanism and colored by energy. The color scale is represented in the figure. Green colored structures represent lower activation barriers and pink colored structures represent high activation barriers.

In both mechanism A and B, the structures in which the catalytic water molecule was closer to the substrate scissile bond ($d1$), are green colored, as expected. Looking only to the mechanism A structures, the orientation of the Asp25A, influenced the obtained barrier. A good orientation of the Thr26B side chain also seemed to be fundamental to the reaction. In this case, if this residue was well oriented to Asp25A, the obtained barriers were higher ($d6$). This could be explained by a higher stabilization of the negatively charged carboxylic group, and, consequently a less capacity to extract a proton from the water molecule (decrease of the pKa), as comment above. The same tendency was observed for the structural water molecule ($d5$).

Concerning the structures that react by mechanism B, small energies (green represented) structures were associated with small values of $d1, d2$ and $d4$ and high values of $d3$. This Figure helps to understand the results presented in **Figure 8.8.**

8.6 Conclusions

The first aim of this work was to understand the effect of conformational fluctuations in the activation barriers of the enzymatic reaction catalyzed by HIV-1 PR.

We collected 30 different structures, equally spaced in time, from a MD simulation and we used them as initial reactants for QM/MM calculation, to study the reaction path of the first step of this mechanism.

The results showed that, the different conformations of the reactant state are associated with two different (albeit similar) chemical mechanisms, with different barriers (ranged from 17.3 to 32.2 kcal/mol). We studied the reasons behind such differences and we identified important interactions in the HIV-1 PR active site that are associated with the progress of the reaction through mechanism A or mechanism B.

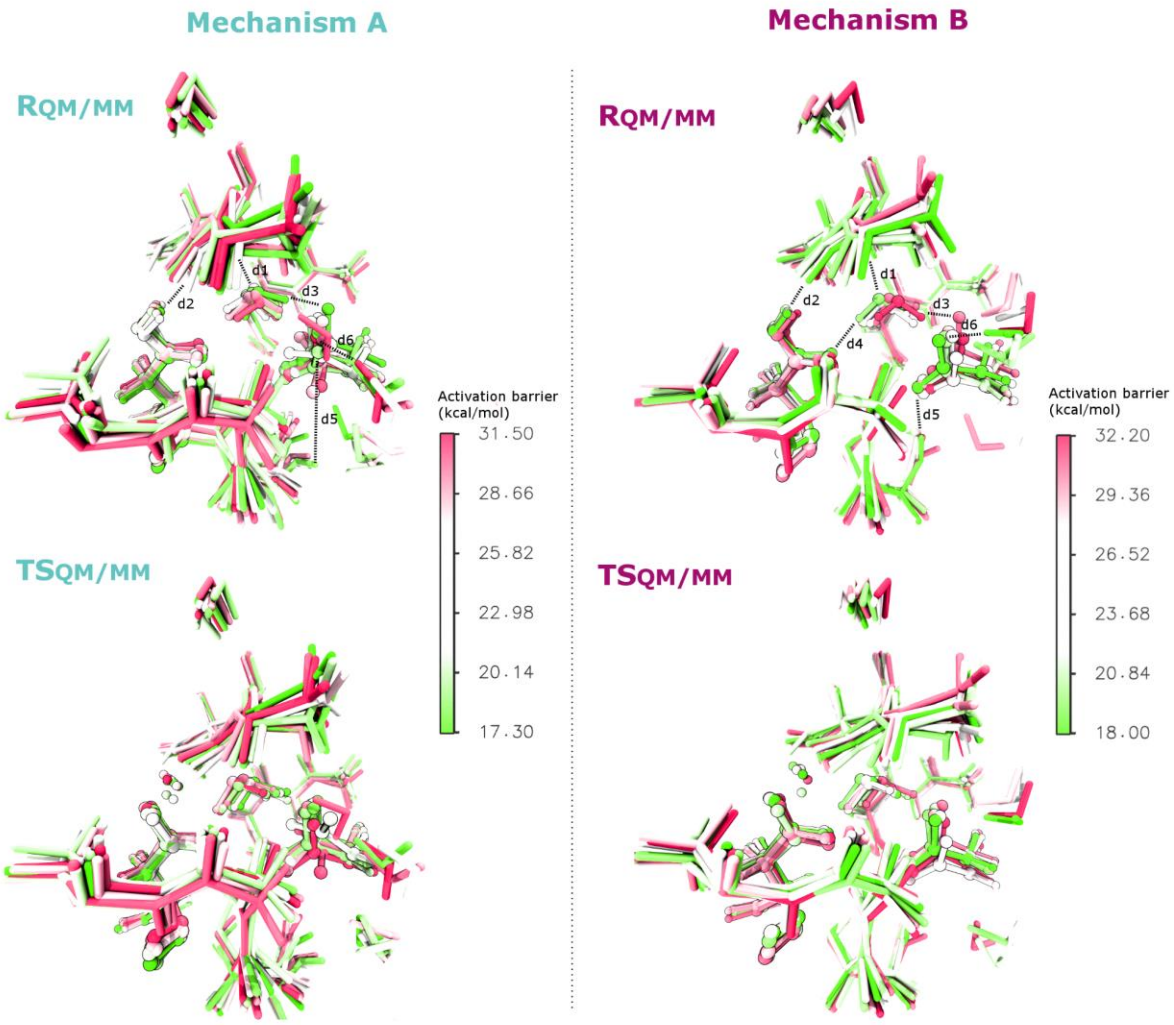


Figure 8.10 Superimposition of the Reactant and Transition state structures at the B3LYP/6-31G(d):ff99SB for mechanism A (first column) and mechanism B (second column). The structures are colored by energy, being the green ones associated with the lower barriers, and the pink ones associated with the higher barriers. For both mechanisms A and B, the range of the energy is similar. Important active site distances were highlighted.

Mechanism A was well described, and accepted, in previous works of HIV-1 PR and other proteases. Mechanism B was considered as led probable, due to the large barriers founded by other authors. Here we observed that, a number of very specific conformations of the HIV-1 PR, also favor the occurrence of the mechanism B, with low activation barriers. In summary we found that mechanism B occurs when the Asp25A is well stabilized through hydrogen bonds with a second structural water molecule and Thr26B, that lower its pKa and make it less suitable to act as a base. The role of this second water molecule was not documented before, even though its prevalence during the MD simulation is consistent with its relevance for the HIV-1 PR mechanism. We show that the positioning of this water molecule could change the progress of the catalytic mechanism, as well as the activation barriers of the reaction. The occurrence of mechanism A with

small barriers was also de.pendent on the correct formation of a set of specific interactions. Weaker hydrogen bonds with Asp25A, which means large distances between the structural water molecule and/or the Thr26B residue and the Asp25A carboxylic group, increased the capacity of this residue to act as a base and, consequently, the progress of the reaction by mechanism, A.

The observed conformational fluctuations occur at the nanosecond timescale (or even faster), which is very fast compared to the turnover of HIV-1 PR.

We believe that our results can help to a better understand, not only of the catalytic mechanism of HIV-1 PR, but also of enzymatic catalysis in general. As in the previous work with α-amylase, we show again, that the conformational fluctuations of enzymes seem to influence the catalysis.

8.7 Supporting Information

Table 8.3 (SI) Activation energies of the QM layer atoms in vacuum ($\Delta E^{\ddagger}_{QM}$), compared to the ONIOM activation energies $\Delta E^{\ddagger}$, without ZPE corrections. The values were calculated with the M06-2X/6-311++G(2d,2p):ff99SB.

Time (ns)	Type of mechanism	$\Delta E^{\ddagger}$ (M06-2X(6-311++G(2d,2p)):ff99SB) kcal/mol	$\Delta E^{\ddagger}_{QM}$ (M06-2X(6-311++G(2d,2p)):ff99SB) kcal/mol
100	B	27.5	26.7
101	A	21.1	26.4
103	B	25.1	22.3
105	A	23.2	23.2
107	A	29.3	22.3
108	B	23.9	22.9
109	A	29.2	23.6
110	A	19.9	18.4
111	B	20.0	16.3
112	B	32.3	29.3
113	A	23.6	20.6
116	A	21.6	24.3
117	A	25.5	22.4
118	A	28.6	24.4
120	A	20.9	25.0
121	A	34.7	35.3
124	B	28.3	22.9
125	B	27.6	21.7
129	B	35.1	27.1

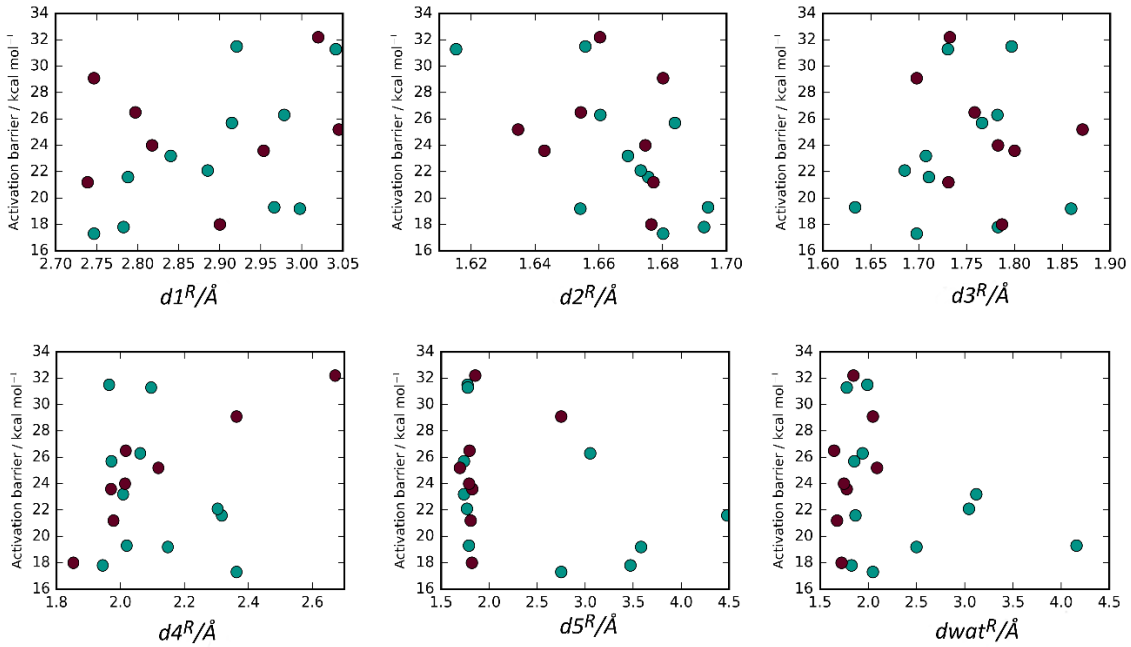


Figure 8.11 (SI) Correlations between six selected active site distances from reactant structures and the corresponding activation barriers. Blue and purple circles correspond to structures that react through mechanism A and B, respectively. There was no evident correlation between these distances and the barriers. However, for mechanism A (blue) a small trend between d1 and the activation barrier can be seen. The smallest distances were associated with small energies and large distances with large energies. The opposite trend can be observed for d2, and the same tendency for d3. Mechanism B seemed to be more dependent on d4. Small values of d4 and d5 seemed to be required for this last mechanism.

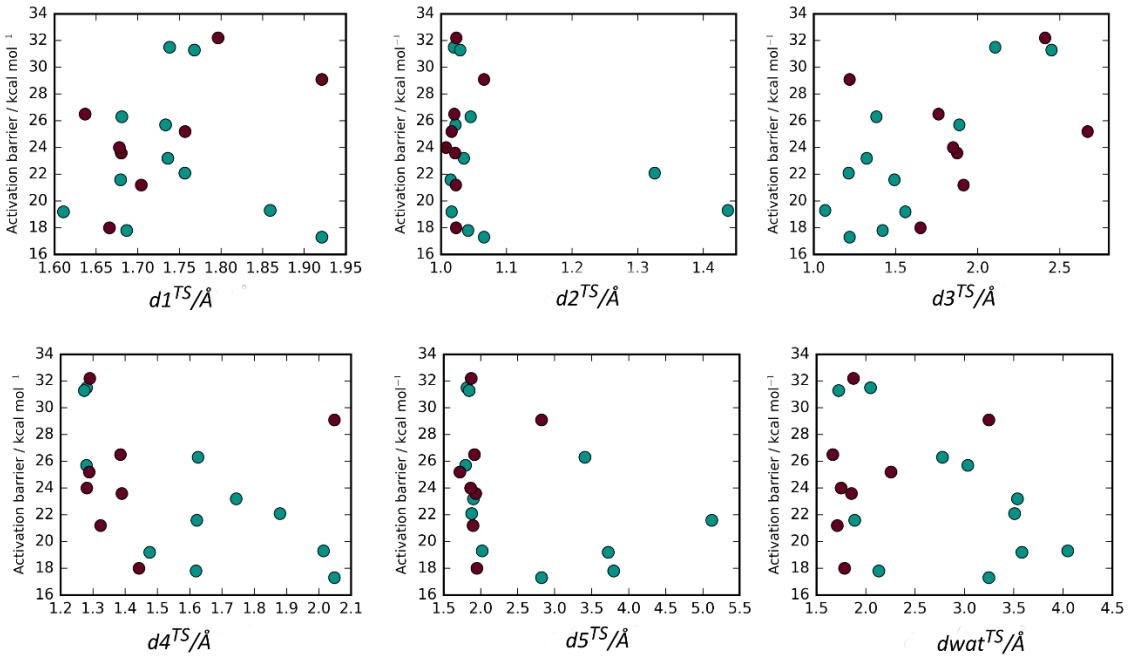


Figure 8.12 (SI) Correlations between six selected active site distances from transition state structures and the corresponding activation barriers. Blue and purple circles correspond to structures that react through mechanism A and B, respectively. Once again, there was no evident correlation between these distances and the barriers. It was visible a small tendency for d1. In general, the small values of this distance were associated with the smallest energy for both mechanisms A and B. Mechanism A was associated with small values of d3 and mechanism B with small values of d4. Mechanism B was also dependent on small values of d5 and dwat.

CHAPTER 9. Conclusion

Understand the origin of the catalytic power of enzymes is one of the most fundamental questions of biochemistry. There are not only fundamental, but also practical interests in finding what makes these biomolecules so efficient. Many proposals have been suggested to account for the catalytic power of enzymes. However, all of them are undemonstrated and this question stills open. The main objective of this thesis was to apply a combination of theoretical and computational methods to produce some insights to help to understand this controversial field.

Different works were presented regarding this objective: a) a general enzyme database was constructed, as planned, with structural and mechanistic information, which were related with enzymatic parameters, to identify patterns responsible for enzyme efficiency and put forward some possible hypothesis for the origin of the catalytic power of enzymes. The main results of this work showed that activation free energy, substrate binding and enzyme efficiency fall in a very narrow range of values for all enzyme classes; b) two different catalytic mechanisms were described, with atomistic detail, for two different enzymes: Human Renin and PatGmacrocyclase; c) a methodologic study showed that the influence of fixing residues during the QM/MM studies is very small. This studied facilitated the preparation of the next QM/MM models; e) the catalytic mechanism of two different enzymes (alpha-amylase and HIV-1 protease) was studied, starting from different initial structures taken from molecular dynamics simulations, to account for the influence of enzyme flexibility on catalysis. These studies showed that the organization of the active site environment is very important for the enzyme catalysis. With these results we can also infer that conformational fluctuations may have an important role on the origin of the catalytic power of enzymes.

After this work, the main question of this thesis remains unanswered, nevertheless we believe that, despite their differences, these studies can give some different and important lights into enzyme catalysis field.

References

(1) Wolfenden, R.; Snider, M. J. The depth of chemical time and the power of enzymes as catalysts. *Acc Chem Res* **2001**, *34*, 938-945.

(2) Hammes, G. G.; Benkovic, S. J.; Hammes-Schiffer, S. Flexibility, Diversity, and Cooperativity: Pillars of Enzyme Catalysis. *Biochemistry-Us* **2011**, *50*, 10422-10430.

(3) Cox, M. M.; Nelson, D. L.: *Lehninger principles of biochemistry*; WH Freeman, 2008.

(4) Berg, J. M.; Tymoczko, J.; Gatto Jr, G. Stryer: Biochemistry. *WH Freeman and Company* **2002**, *5*, 306-307.

(5) Pasteur, L. Mémoire sur la fermentation appelée lactique (Extrait par l'auteur). *Molecular Medicine* **1995**, *1*, 599.

(6) Kühne, W. Über das Verhalten verschiedener organisirter und sog. ungeformter Fermente, Separat-Abdruck aus den Verhandlungen des Heidelb. Naturhist.-Med. Vereins. NS 13.[Reprinted 1976. *FEBS Lett* **1876**, *62*, E3-E12.

(7) Sumner, J. B. The isolation and crystallization of the enzyme urease preliminary paper. *Journal of Biological Chemistry* **1926**, *69*, 435-441.

(8) Bernal, J. D.; Crowfoot, D. X-ray photographs of crystalline pepsin. *Nature* **1934**, *133*, 794.

(9) Kunitz, M.; Northrop, J. H. Crystalline chymo-trypsin and chymo-trypsinogen: I. Isolation, crystallization, and general properties of a new proteolytic enzyme and its precursor. *The Journal of general physiology* **1935**, *18*, 433-458.

(10) Haldane, J. S. 1930. Enzymes. *London, Longmann and Green* **1930**.

(11) Pauling, L. Molecular architecture and biological reactions. *Chemical and engineering news* **1946**, *24*, 1375-1377.

(12) Wolfenden, R. Benchmark reaction rates, the stability of biological molecules in water, and the evolution of catalytic power in enzymes. *Annu Rev Biochem* **2011**, *80*, 645-667.

(13) Benkovic, S. J.; Hammes-Schiffer, S. A perspective on enzyme catalysis. *Science* **2003**, *301*, 1196-1202.

(14) Knowles, J. R. Enzyme catalysis: not different, just better. *Nature* **1991**, *350*, 121.

(15) Cannon, W. R.; Singleton, S. F.; Benkovic, S. J. A perspective on biological catalysis. *Nature Structural and Molecular Biology* **1996**, *3*, 821.

(16) Jencks, W. P.: *Binding energy, specificity, and enzymic catalysis: the circe effect*; Wiley Online Library, 1975.

(17) Sadiq, S. K.; Coveney, P. V. Computing the role of near attack conformations in an enzyme-catalyzed nucleophilic bimolecular reaction. *J Chem Theory Comput* **2015**, *11*, 316-324.

(18) Warshel, A. Energetics of enzyme catalysis. *Proceedings of the National Academy of Sciences* **1978**, *75*, 5250-5254.

(19) Warshel, A. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *Journal of Biological Chemistry* **1998**, *273*, 27035-27038.

(20) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. Electrostatic basis for enzyme catalysis. *Chemical reviews* **2006**, *106*, 3210-3235.

(21) Anderson, J. B. Predicting rare events in molecular dynamics. *Advances in Chemical Physics* **1995**, *91*, 381-432.

(22) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. Current status of transition-state theory. *The Journal of physical chemistry* **1996**, *100*, 12771-12800.

(23) Hammes-Schiffer, S.; Benkovic, S. J. Relating protein motion to catalysis. *Annu Rev Biochem* **2006**, *75*, 519-541.

(24) Mulholland, A. J. Dispelling the effects of a sorceress in enzyme catalysis. *Proc Natl Acad Sci U S A* **2016**, *113*, 2328-2330.

(25) Kamerlin, S. C. L.; Warshel, A. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins* **2010**, *78*, 1339-1375.

(26) Giraldo, J.; Roche, D.; Rovira, X.; Serra, J. The catalytic power of enzymes: conformational selection or transition state stabilization? *FEBS Lett* **2006**, *580*, 2170-2177.

(27) Zhang, X. Y.; Houk, K. N. Why enzymes are proficient catalysts: Beyond the Pauling paradigm. *Accounts Chem Res* **2005**, *38*, 379-385.

(28) Cannon, W. R.; Benkovic, S. J. Solvation, reorganization energy, and biological catalysis. *J Biol Chem* **1998**, *273*, 26257-26260.

(29) Cohen, S. G.; Vaidya, V. M.; Schultz, R. M. Active Site of alpha-Chymotrypsin Activation by Association-Desolvation. *Proc Natl Acad Sci U S A* **1970**, *66*, 249-256.

(30) Devi-Kesavan, L. S.; Gao, J. Combined QM/MM study of the mechanism and kinetic isotope effect of the nucleophilic substitution reaction in haloalkane dehalogenase. *J Am Chem Soc* **2003**, *125*, 1532-1540.

(31) Dewar, M. J.; Storch, D. M. Alternative view of enzyme reactions. *Proc Natl Acad Sci U S A* **1985**, *82*, 2225-2229.

(32) Lee, J. K.; Houk, K. N. A proficient enzyme revisited: the predicted mechanism for orotidine monophosphate decarboxylase. *Science* **1997**, *276*, 942-945.

(33) Warshel, A. Perspective on "The energetics of enzymatic reactions" - Warshel A (1978) Proc Natl Acad Sci USA 75 : 5250. *Theor Chem Acc* **2000**, *103*, 337-339.

(34) Warshel, A.; Florian, J. Computer simulations of enzyme catalysis: Finding out what has been optimized by evolution. *P Natl Acad Sci USA* **1998**, *95*, 5950-5955.

(35) Feierberg, I.; Aqvist, J. Computational modeling of enzymatic keto-enol isomerization reactions. *Theor Chem Acc* **2002**, *108*, 71-84.

(36) Hanoian, P.; Liu, C. T.; Hammes-Schiffer, S.; Benkovic, S. Perspectives on electrostatics and conformational motions in enzyme catalysis. *Acc Chem Res* **2015**, *48*, 482-489.

(37) Zhou, H. X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem Rev* **2018**, *118*, 1691-1741.

(38) Sines, J. J.; Allison, S. A.; Mccammon, J. A. Point-Charge Distributions and Electrostatic Steering in Enzyme Substrate Encounter - Brownian Dynamics of Modified Copper-Zinc Superoxide Dismutases. *Biochemistry-Us* **1990**, *29*, 9403-9412.

(39) Kazemi, M.; Himo, F.; Aqvist, J. Enzyme catalysis by entropy without Circe effect. *P Natl Acad Sci USA* **2016**, *113*, 2406-2411.

(40) Page, M. I.; Jencks, W. P. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc Natl Acad Sci U S A* **1971**, *68*, 1678-1683.

(41) Aqvist, J.; Kazemi, M.; Isaksen, G. V.; Brandsdal, B. O. Entropy and Enzyme Catalysis. *Accounts Chem Res* **2017**, *50*, 199-207.

(42) Stanton, R. V.; Perakyla, M.; Bakowies, D.; Kollman, P. A. Combined ab initio and free energy calculations to study reactions in enzymes and solution: Amide hydrolysis in trypsin and aqueous solution. *Journal of the American Chemical Society* **1998**, *120*, 3448-3457.

(43) Piazzetta, P.; Marino, T.; Russo, N. Mechanistic Explanation of the Weak Carbonic Anhydrase's Esterase Activity. *Molecules* **2017**, *22*.

(44) Anderson, V. E. Quantifying energetic contributions to ground state destabilization. *Arch Biochem Biophys* **2005**, *433*, 27-33.

(45) Ruben, E. A.; Schwans, J. P.; Sonnett, M.; Natarajan, A.; Gonzalez, A.; Tsai, Y. S.; Herschlag, D. Ground State Destabilization from a Positioned General Base in the Ketosteroid Isomerase Active Site. *Biochemistry-Us* **2013**, *52*, 1074-1081.

(46) Phillips, R. S.; Vita, A.; Spivey, J. B.; Rudloff, A. P.; Driscoll, M. D.; Hay, S. Ground-State Destabilization by Phe-448 and Phe-449 Contributes to Tyrosine Phenol-Lyase Catalysis. *Acs Catal* **2016**, *6*, 6770-6779.

(47) Robinson, R. Don't Get Too Comfortable: Destabilizing the Ground State to Speed a Reaction. *Plos Biol* **2013**, *11*.

(48) Dafforn, A.; Koshland, D. E., Jr. Theoretical aspects of orbital steering. *Proc Natl Acad Sci U S A* **1971**, *68*, 2463-2467.

(49) Mesecar, A. D.; Stoddard, B. L.; Koshland, D. E., Jr. Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences. *Science* **1997**, *277*, 202-206.

(50) Cleland, W. W.; Kreevoy, M. M. Low-barrier hydrogen bonds and enzymic catalysis. *Science* **1994**, *264*, 1887-1890.

(51) Cleland, W. W.; Frey, P. A.; Gerlt, J. A. The low barrier hydrogen bond in enzymatic catalysis. *J Biol Chem* **1998**, *273*, 25529-25532.

(52) Ishida, T. Low-barrier hydrogen bond hypothesis in the catalytic triad residue of serine proteases: correlation between structural rearrangement and chemical shifts in the acylation process. *Biochemistry-Us* **2006**, *45*, 5413-5420.

(53) Warshel, A.; Papazyan, A. Energy considerations show that low-barrier hydrogen bonds do not offer a catalytic advantage over ordinary hydrogen bonds. *Proceedings of the National Academy of Sciences* **1996**, *93*, 13665-13670.

(54) Warshel, A.; Papazyan, A.; Kollman, P. A. On low-barrier hydrogen bonds and enzyme catalysis. *Science* **1995**, *269*, 102-106.

(55) Hur, S.; Bruice, T. C. The near attack conformation approach to the study of the chorismate to prephenate reaction. *Proceedings of the National Academy of Sciences* **2003**, *100*, 12015-12020.

(56) Hur, S.; Bruice, T. C. The mechanism of catalysis of the chorismate to prephenate reaction by the Escherichia coli mutase enzyme. *Proceedings of the National Academy of Sciences* **2002**, *99*, 1176-1181.

(57) Ranaghan, K. E.; Mulholland, A. J. Conformational effects in enzyme catalysis: QM/MM free energy calculation of the 'NAC'contribution in chorismate mutase. *Chemical communications* **2004**, 1238-1239.

(58) Guo, H.; Cui, Q.; Lipscomb, W. N.; Karplus, M. Understanding the role of active-site residues in chorismate mutase catalysis from molecular-dynamics simulations. *Angewandte Chemie International Edition* **2003**, *42*, 1508-1511.

(59) Kohen, A. Role of dynamics in enzyme catalysis: substantial versus semantic controversies. *Acc Chem Res* **2015**, *48*, 466-473.

(60) Klinman, J. P. Dynamically achieved active site precision in enzyme catalysis. *Acc Chem Res* **2015**, *48*, 449-456.

(61) Callender, R.; Dyer, R. B. The dynamical nature of enzymatic catalysis. *Acc Chem Res* **2015**, *48*, 407-413.

(62) Agarwal, P. K.; Billeter, S. R.; Rajagopalan, P. T.; Benkovic, S. J.; Hammes-Schiffer, S. Network of coupled promoting motions in enzyme catalysis. *Proc Natl Acad Sci U S A* **2002**, *99*, 2794-2799.

(63) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **2007**, *450*, 913-916.

(64) Piana, S.; Bucher, D.; Carloni, P.; Rothlisberger, U. Reaction mechanism of HIV-1 protease by hybrid carparrinello/classical MD simulations. *J Phys Chem B* **2004**, *108*, 11139-11149.

(65) Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964-972.

(66) Kurplus, M.; McCammon, J. Dynamics of proteins: elements and function. *Annual review of biochemistry* **1983**, *52*, 263-300.

(67) Daniel, R. M.; Dunn, R. V.; Finney, J. L.; Smith, J. C. The role of dynamics in enzyme activity. *Annu Rev Biophys Biomol Struct* **2003**, *32*, 69-92.

(68) Roca, M.; Andres, J.; Moliner, V.; Tunon, I.; Bertran, J. On the nature of the transition state in catechol O-methyltransferase. A complementary study based on molecular dynamics and potential energy surface explorations. *J Am Chem Soc* **2005**, *127*, 10648-10655.

(69) Ruiz-Pernia, J. J.; Tunon, I.; Moliner, V.; Hynes, J. T.; Roca, M. Dynamic effects on reaction rates in a Michael addition catalyzed by chalcone isomerase. Beyond the frozen environment approach. *J Am Chem Soc* **2008**, *130*, 7477-7488.

(70) Garcia-Meseguer, R.; Marti, S.; Ruiz-Pernia, J. J.; Moliner, V.; Tunon, I. Studying the role of protein dynamics in an SN2 enzyme reaction using free-energy surfaces and solvent coordinates. *Nat Chem* **2013**, *5*, 566-571.

(71) Ribeiro, A. n. J.; Santos-Martins, D.; Russo, N.; Ramos, M. J.; Fernandes, P. A. Enzymatic flexibility and reaction rate: A QM/MM study of HIV-1 protease. *Acs Catal* **2015**, *5*, 5617-5626.

(72) Agarwal, P. K.; Billeter, S. R.; Hammes-Schiffer, S. Nuclear quantum effects and enzyme dynamics in dihydrofolate reductase catalysis. *The Journal of Physical Chemistry B* **2002**, *106*, 3283-3293.

(73) Radkiewicz, J. L.; Brooks, C. L. Protein dynamics in enzymatic catalysis: exploration of dihydrofolate reductase. *Journal of the American Chemical Society* **2000**, *122*, 225-231.

(74) Adamczyk, A. J.; Cao, J.; Kamerlin, S. C.; Warshel, A. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc Natl Acad Sci U S A* **2011**, *108*, 14115-14120.

(75) Luk, L. Y.; Javier Ruiz-Pernia, J.; Dawson, W. M.; Roca, M.; Loveridge, E. J.; Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J.; Tunon, I.; Moliner, V.; Allemann, R. K. Unraveling the role of protein dynamics in dihydrofolate reductase catalysis. *Proc Natl Acad Sci U S A* **2013**, *110*, 16344-16349.

(76) Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *science* **2006**, *313*, 1638-1642.

(77) Caratzoulas, S.; Mincer, J. S.; Schwartz, S. D. Identification of a protein-promoting vibration in the reaction catalyzed by horse liver alcohol dehydrogenase. *Journal of the American Chemical Society* **2002**, *124*, 3270-3276.

(78) Mincer, J. S.; Schwartz, S. D. A computational method to identify residues important in creating a protein promoting vibration in enzymes. *The Journal of Physical Chemistry B* **2003**, *107*, 366-371.

(79) Pisliakov, A. V.; Cao, J.; Kamerlin, S. C.; Warshel, A. Enzyme millisecond conformational dynamics do not catalyze the chemical step. *Proceedings of the National Academy of Sciences* **2009**, *106*, 17359-17364.

(80) Olsson, M. H.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions. *Journal of chemical theory and computation* **2011**, *7*, 525-537.

(81) Olsson, M. H. M.; Parson, W. W.; Warshel, A. Dynamical contributions to enzyme catalysis: Critical tests of a popular hypothesis. *Chemical Reviews* **2006**, *106*, 1737-1756.

(82) Liang, Z.-X.; Klinman, J. P. Structural bases of hydrogen tunneling in enzymes: progress and puzzles. *Current opinion in structural biology* **2004**, *14*, 648-655.

(83) Kohen, A.; Klinman, J. P. Hydrogen tunneling in biology. *Chemistry & biology* **1999**, *6*, R191-198.

(84) Sutcliffe, M. J.; Scrutton, N. S. Enzyme catalysis: over-the-barrier or through-the-barrier? *Trends Biochem Sci* **2000**, *25*, 405-408.

(85) Sikorski, R. S.; Wang, L.; Markham, K. A.; Rajagopalan, P. T.; Benkovic, S. J.; Kohen, A. Tunneling and coupled motion in the Escherichia coli dihydrofolate reductase catalysis. *J Am Chem Soc* **2004**, *126*, 4778-4779.

(86) Maglia, G.; Allemann, R. K. Evidence for environmentally coupled hydrogen tunneling during dihydrofolate reductase catalysis. *J Am Chem Soc* **2003**, *125*, 13372-13373.

(87) Pang, J.; Pu, J.; Gao, J.; Truhlar, D. G.; Allemann, R. K. Hydride transfer reaction catalyzed by hyperthermophilic dihydrofolate reductase is dominated by quantum mechanical tunneling and is promoted by both inter- and intramonomeric correlated motions. *J Am Chem Soc* **2006**, *128*, 8015-8023.

(88) Sousa, S. F.; Ribeiro, A. J.; Neves, R. P.; Brás, N. F.; Cerqueira, N. M.; Fernandes, P. A.; Ramos, M. J. Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2017**, *7*.

(89) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Computational enzymatic catalysis - clarifying enzymatic mechanisms with the help of computers. *Phys Chem Chem Phys* **2012**, *14*, 12431-12441.

(90) Vlachakis, D.; Bencurova, E.; Papangelopoulos, N.; Kossida, S. Current state-of-the-art molecular dynamics methods and applications. *Adv Protein Chem Struct Biol* **2014**, *94*, 269-313.

(91) Cramer, C. J.: *Essentials of computational chemistry: theories and models*; John Wiley & Sons, 2013.

(92) Young, D.: *Computational chemistry: a practical guide for applying techniques to real world problems*; John Wiley & Sons, 2004.

(93) Neese, F.; Bredow, T.; Wennmohs, F. Introduction to Computational Chemistry. **2007**.

(94) Cerqueira, N.; Fernandes, P.; Ramos, M. J. Protocol for Computational Enzymatic Reactivity Based on Geometry Optimisation. *ChemPhysChem* **2018**, *19*, 669-689.

(95) van der Kamp, M. W.; Mulholland, A. J. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry-Us* **2013**, *52*, 2708-2728.

(96) Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J. Computational enzymology. *Chemical Communications* **2010**, *46*, 2354-2372.

(97) Jensen, F.: *Introduction to computational chemistry*; John wiley & sons, 2017.

(98) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **1995**, *117*, 5179-5197.

(99) Debye, P. Näherungsformeln für die Zylinderfunktionen für große Werte des Arguments und unbeschränkt veränderliche Werte des Index. *Mathematische Annalen* **1909**, *67*, 535-558.

(100) Hestenes, M. R.; Stiefel, E.: *Methods of conjugate gradients for solving linear systems*; NBS Washington, DC, 1952; Vol. 49.

(101) Shewchuk, J. R.: An introduction to the conjugate gradient method without the agonizing pain. Carnegie-Mellon University. Department of Computer Science, 1994.

(102) González, M. Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique* **2011**, *12*, 169-200.

(103) D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C.Walker,W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, K.F.Wong, F. Paesani, J. Vanicek, R.M.Wolf, J. Liu, X.Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J.Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman: AMBER 12. University of California, San Francisco, 2012.

(104) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977**, *23*, 327-341.

(105) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **1983**, *79*, 926-935.

(106) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *The Journal of chemical physics* **1993**, *98*, 10089-10092.

(107) Ewald, P. P. The calculation of optical and electrostatic grid potential. *Ann. Phys* **1921**, *64*, 253-287.

(108) Langevin, P. Sur la théorie du mouvement brownien. *Compt. Rendus* **1908**, *146*, 530-533.

(109) Berendsen, H. J.; Postma, J. v.; van Gunsteren, W. F.; DiNola, A.; Haak, J. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **1984**, *81*, 3684-3690.

(110) Case, D.; Darden, T.; Cheatham III, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K. AMBER 12; University of California: San Francisco, 2012. *There is no corresponding record for this reference* **2010**, 1-826.

(111) Born, M.; Oppenheimer, R. Zur quantentheorie der molekeln. *Annalen der physik* **1927**, *389*, 457-484.

(112) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A* **1988**, *38*, 3098.

(113) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Physical review letters* **1996**, *77*, 3865.

(114) Becke, A. D. Density-functional thermochemistry. IV. A new dynamical correlation functional and implications for exact-exchange mixing. *The Journal of chemical physics* **1996**, *104*, 1040-1046.

(115) Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Physical Review B* **1988**, *37*, 785-789.

(116) Siegbahn, P. E. The performance of hybrid DFT for mechanisms involving transition metal complexes in enzymes. *JBIC Journal of Biological Inorganic Chemistry* **2006**, *11*, 695-701.

(117) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical Review B* **1992**, *46*, 6671.

(118) Adamo, C.; Barone, V. Toward reliable adiabatic connection models free from adjustable parameters. *Chemical Physics Letters* **1997**, *274*, 242-250.

(119) Adamo, C.; Barone, V. Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The m PW and m PW1PW models. *The Journal of chemical physics* **1998**, *108*, 664-675.

(120) Zhao, Y.; Truhlar, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *The Journal of chemical physics* **2006**, *125*, 194101.

(121) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor Chem Acc* **2008**, *120*, 215-241.

(122) Fouda, A.; Ryde, U. Does the DFT Self-Interaction Error Affect Energies Calculated in Proteins with Large QM Systems? *Journal of chemical theory and computation* **2016**, *12*, 5667-5679.

(123) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. 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* **2010**, *132*, 154104.

(124) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* **2011**, *32*, 1456-1465.

(125) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods .9. Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *Journal of Chemical Physics* **1971**, *54*, 724-+.

(126) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of molecular biology* **1976**, *103*, 227-249.

(127) Lin, H.; Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theor Chem Acc* **2007**, *117*, 185.

(128) Vreven, T.; Morokuma, K. Hybrid methods: Oniom (qm: mm) and qm/mm. *Annual reports in computational chemistry* **2006**, *2*, 35-51.

(129) Hirao, H.; Xu, K.; Chuanprasit, P.; Moeljadi, A. M. P.; Morokuma, K. Key Concepts and Applications of ONIOM Methods. *Simulating Enzyme Reactivity: Computational Methods in Enzyme Catalysis* **2016**, 245.

(130) Quesne, M. G.; Borowski, T.; de Visser, S. P. Quantum mechanics/molecular mechanics modeling of enzymatic processes: caveats and breakthroughs. *Chem-Eur J* **2016**, *22*, 2562-2581.

(131) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition* **2009**, *48*, 1198-1229.

(132) Maseras, F.; Morokuma, K. IMOMM: A new integrated ab initio+ molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J Comput Chem* **1995**, *16*, 1170-1179.

(133) Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints. *J Comput Chem* **2003**, *24*, 760-769.

(134) Zhou, J.; Tao, P.; Fisher, J. F.; Shi, Q.; Mobashery, S.; Schlegel, H. B. QM/MM studies of the matrix metalloproteinase 2 (MMP2) inhibition mechanism of (S)-SB-3CT and its oxirane analogue. *Journal of chemical theory and computation* **2010**, *6*, 3580-3587.

(135) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. 09, Revision D. 01, Gaussian. *Inc., Wallingford, CT* **2009**.

(136) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res* **2000**, *28*, 235-242.

(137) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Structure, Function, and Bioinformatics* **2005**, *61*, 704-721.

(138) Bas, D. C.; Rogers, D. M.; Jensen, J. H. Very fast prediction and rationalization of pKa values for protein–ligand complexes. *Proteins: Structure, Function, and Bioinformatics* **2008**, *73*, 765-783.

(139) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: a server for estimating p K as and adding missing hydrogens to macromolecules. *Nucleic Acids Res* **2005**, *33*, W368-W371.

(140) Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys Chem Chem Phys* **2011**, *13*, 6670-6688.

(141) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712-725.

(142) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1950-1958.

(143) Ramos, M. J.; Fernandes, P. A. Computational enzymatic catalysis. *Accounts Chem Res* **2008**, *41*, 689-698.

(144) Ochterski, J. W. Thermochemistry in gaussian. *Gaussian Inc* **2000**, 1-19.

(145) McQuarrie, D. A.; Simon, J. D.: *Molecular thermodynamics*; University Science Books Sausalito, CA, 1999.

(146) Sgrignani, J.; Magistrato, A. QM/MM MD simulations on the enzymatic pathway of the human flap endonuclease (hFEN1) elucidating common cleavage pathways to RNase H enzymes. *Acs Catal* **2015**, *5*, 3864-3875.

(147) Senn, H. M.; Thiel, S.; Thiel, W. Enzymatic hydroxylation in p-hydroxybenzoate hydroxylase: a case study for QM/MM molecular dynamics. *Journal of chemical theory and computation* **2005**, *1*, 494-505.

(148) Ruggiero, G. D.; Williams, I. H.; Roca, M.; Moliner, V.; Tuñón, I. QM/MM determination of kinetic isotope effects for COMT-catalyzed methyl transfer does not support compression hypothesis. *Journal of the American Chemical Society* **2004**, *126*, 8634-8635.

(149) Car, R.; Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Physical review letters* **1985**, *55*, 2471.

(150) Warshel, A.; Weiss, R. M. An Empirical Valence Bond Approach for Comparing Reactions in Solutions and in Enzymes. *J Am Chem Soc* **1980**, *102*, 6218-6226.

(151) Warshel, A.; Weiss, R. Empirical valence bond calculations of enzyme catalysis. *Annals of the New York Academy of Sciences* **1981**, *367*, 370-382.

(152) Kamerlin, S. C. L.; Warshel, A. The empirical valence bond model: theory and applications. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 30-45.

(153) Siegbahn, P. E.; Himo, F. The quantum chemical cluster approach for modeling enzyme reactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 323-336.

(154) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. How enzymes work: Analysis by modern rate theory and computer simulations. *Science* **2004**, *303*, 186-195.

(155) Pauling, L. Nature of Forces between Large Molecules of Biological Interest. *Nature* **1948**, *161*, 707-709.

(156) Villa, J.; Strajbl, M.; Glennon, T. M.; Sham, Y. Y.; Chu, Z. T.; Warshel, A. How important are entropic contributions to enzyme catalysis? *P Natl Acad Sci USA* **2000**, *97*, 11899-11904.

(157) Menger, F. M.; Glass, L. E. Contribution of Orbital Alignment to Organic and Enzymatic Reactivity. *Journal of the American Chemical Society* **1980**, *102*, 5404-5406.

(158) Cleland, W. W. The low-barrier hydrogen bond in enzymic catalysis. *Adv Phys Org Chem* **2010**, *44*, 1-17.

(159) Tunon, I.; Laage, D.; Hynes, J. T. Are there dynamical effects in enzyme catalysis? Some thoughts concerning the enzymatic chemical step. *Arch Biochem Biophys* **2015**, *582*, 42-55.

(160) Bruice, T. C.; Benkovic, S. J. Chemical basis for enzyme catalysis. *Biochemistry-Us* **2000**, *39*, 6267-6274.

(161) Gutteridge, A.; Thornton, J. M. Understanding nature's catalytic toolkit. *Trends in Biochemical Sciences* **2005**, *30*, 622-629.

(162) Hollfelder, F.; Kirby, A. J.; Tawfik, D. S. On the magnitude and specificity of medium effects in enzyme-like catalysts for proton transfer. *J Org Chem* **2001**, *66*, 5866-5874.

(163) Agarwal, P. K. Role of protein dynamics in reaction rate enhancement by enzymes. *Journal of the American Chemical Society* **2005**, *127*, 15248-15256.

(164) Sousa, S. F.; Ramos, M. J.; Lim, C.; Fernandes, P. A. Relationship between Enzyme/Substrate Properties and Enzyme Efficiency in Hydrolases. *Acs Catal* **2015**, *5*, 5877-5887.

(165) Schomburg, I.; Chang, A.; Schomburg, D. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res* **2002**, *30*, 47-49.

(166) Schomburg, I.; Chang, A.; Placzek, S.; Sohngen, C.; Rother, M.; Lang, M.; Munaretto, C.; Ulas, S.; Stelzer, M.; Grote, A.; Scheer, M.; Schomburg, D. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res* **2013**, *41*, D764-D772.

(167) Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **2003**, *31*, 3784-3788.

(168) Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; de Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; Grosdidier, A.; Hernandez, C.; Ioannidis, V.; Kuznetsov, D.; Liechti, R.; Moretti, S.; Mostaguir, K.; Redaschi, N.; Rossier, G.; Xenarios, I.; Stockinger, H. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* **2012**, *40*, W597-W603.

(169) Bairoch, A.; Boeckmann, B. The Swiss-Prot Protein-Sequence Data-Bank. *Nucleic Acids Res* **1991**, *19*, 2247-2248.

(170) Smith, R. D.; Engdahl, A. L.; Dunbar, J. B.; Carlson, H. A. Biophysical Limits of Protein-Ligand Binding. *J Chem Inf Model* **2012**, *52*, 2098-2106.

(171) Hu, L. G.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A. Binding MOAD (Mother of All Databases). *Proteins* **2005**, *60*, 333-340.

(172) Dudev, T.; Lim, C. Metal binding affinity and selectivity in metalloproteins: insights from computational studies. *Annu Rev Biophys* **2008**, *37*, 97-116.

(173) Dudev, T.; Lim, C. Competition among metal ions for protein binding sites: determinants of metal ion selectivity in proteins. *Chem Rev* **2014**, *114*, 538-556.

(174) Sousa, S. F.; Lopes, A. B.; Fernandes, P. A.; Ramos, M. J. The Zinc proteome: a tale of stability and functionality. *Dalton Trans* **2009**, 7946-7956.

(175) Bezencon, O.; Bur, D.; Weller, T.; Richard-Bildstein, S.; Remen, L.; Sifferlen, T.; Corminboeuf, O.; Grisostomi, C.; Boss, C.; Prade, L.; Delahaye, S.; Treiber, A.; Strickner, P.; Binkert, C.; Hess, P.; Steiner, B.; Fischli, W. Design and preparation of potent, nonpeptidic, bioavailable renin inhibitors. *Journal of medicinal chemistry* **2009**, *52*, 3689-3702.

(176) Blundell, T.; Sibanda, B. L.; Pearl, L. Three-dimensional structure, specificity and catalytic mechanism of renin. *Nature* **1983**, *304*, 273-275.

(177) Eder, J.; Hommel, U.; Cumin, F.; Martoglio, B.; Gerhartz, B. Aspartic proteases in drug discovery. *Current pharmaceutical design* **2007**, *13*, 271-285.

(178) MacGregor, G. A.; Markandu, N. D.; Roulston, J. E.; Jones, J. C.; Morton, J. J. Maintenance of blood pressure by the renin-angiotensin system in normal man. *Nature* **1981**, *291*, 329-331.

(179) Rahuel, J.; Rasetti, V.; Maibaum, J.; Rueger, H.; Goschke, R.; Cohen, N. C.; Stutz, S.; Cumin, F.; Fuhrer, W.; Wood, J. M.; Grutter, M. G. Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of human renin. *Chemistry & biology* **2000**, *7*, 493-504.

(180) Verdecchia, P.; Angeli, F.; Mazzotta, G.; Gentile, G.; Reboldi, G. The renin angiotensin system in the development of cardiovascular disease: role of aliskiren in risk reduction. *Vascular health and risk management* **2008**, *4*, 971-981.

(181) Campbell, D. J. Angiotensin II generation in vivo: does it involve enzymes other than renin and angiotensin-converting enzyme? *Journal of the renin-angiotensin-aldosterone system : JRAAS* **2012**, *13*, 314-316.

(182) Campbell, D. J.; Alexiou, T.; Xiao, H. D.; Fuchs, S.; McKinley, M. J.; Corvol, P.; Bernstein, K. E. Effect of reduced angiotensin-converting enzyme gene expression and angiotensin-converting enzyme inhibition on angiotensin and bradykinin peptide levels in mice. *Hypertension* **2004**, *43*, 854-859.

(183) Dinh, D. T.; Frauman, A. G.; Johnston, C. I.; Fabiani, M. E. Angiotensin receptors: distribution, signalling and function. *Clinical science* **2001**, *100*, 481-492.

(184) Crowley, S. D.; Coffman, T. M. Recent advances involving the renin-angiotensin system. *Experimental cell research* **2012**, *318*, 1049-1056.

(185) Katsurada, A.; Hagiwara, Y.; Miyashita, K.; Satou, R.; Miyata, K.; Ohashi, N.; Navar, L. G.; Kobori, H. Novel sandwich ELISA for human angiotensinogen. *American journal of physiology. Renal physiology* **2007**, *293*, F956-960.

(186) Cumin, F.; Le-Nguyen, D.; Castro, B.; Menard, J.; Corvol, P. Comparative enzymatic studies of human renin acting on pure natural or synthetic substrates. *Biochimica et biophysica acta* **1987**, *913*, 10-19.

(187) Tice, C. M. Renin inhibitors. *Annual Reports in Medicinal Chemistry* **2006**, *41*, 155.

(188) Jensen, C.; Herold, P.; Brunner, H. R. Aliskiren: the first renin inhibitor for clinical treatment. *Nature reviews. Drug discovery* **2008**, *7*, 399-410.

(189) Pilz, B.; Shagdarsuren, E.; Wellner, M.; Fiebeler, A.; Dechend, R.; Gratze, P.; Meiners, S.; Feldman, D. L.; Webb, R. L.; Garrelds, I. M.; Jan Danser, A. H.; Luft, F. C.; Muller, D.

N. Aliskiren, a human renin inhibitor, ameliorates cardiac and renal damage in double-transgenic rats. *Hypertension* **2005**, *46*, 569-576.

(190) Wood, J. M.; Maibaum, J.; Rahuel, J.; Grutter, M. G.; Cohen, N. C.; Rasetti, V.; Ruger, H.; Goschke, R.; Stutz, S.; Fuhrer, W.; Schilling, W.; Rigollier, P.; Yamaguchi, Y.; Cumin, F.; Baum, H. P.; Schnell, C. R.; Herold, P.; Mah, R.; Jensen, C.; O'Brien, E.; Stanton, A.; Bedigian, M. P. Structure-based design of aliskiren, a novel orally effective renin inhibitor. *Biochemical and biophysical research communications* **2003**, *308*, 698-705.

(191) Sielecki, A. R.; Hayakawa, K.; Fujinaga, M.; Murphy, M. E. P.; Fraser, M.; Muir, A. K.; Carilli, C. T.; Lewicki, J. A.; Baxter, J. D.; James, M. N. G. Structure of Recombinant Human Renin, a Target for Cardiovascular-Active Drugs, at 2.5 a Resolution. *Science* **1989**, *243*, 1346-1351.

(192) Gradman, A. H.; Kad, R. Renin inhibition in hypertension. *Journal of the American College of Cardiology* **2008**, *51*, 519-528.

(193) Nakagawa, T.; Akaki, J.; Satou, R.; Takaya, M.; Iwata, H.; Katsurada, A.; Nishiuchi, K.; Ohmura, Y.; Suzuki, F.; Nakamura, Y. The His-Pro-Phe motif of angiotensinogen is a crucial determinant of the substrate specificity of renin. *Biol Chem* **2007**, *388*, 237-246.

(194) Gorfe, A. A.; Caflisch, A. Functional plasticity in the substrate binding site of β-secretase. *Structure* **2005**, *13*, 1487-1498.

(195) Politi, A.; Durdagi, S.; Moutevelis-Minakakis, P.; Kokotos, G.; Papadopoulos, M. G.; Mavromoustakos, T. Application of 3D QSAR CoMFA/CoMSIA and in silico docking studies on novel renin inhibitors against cardiovascular diseases. *Eur J Med Chem* **2009**, *44*, 3703-3711.

(196) Sibanda, B. L.; Blundell, T.; Hobart, P. M.; Fogliano, M.; Bindra, J. S.; Dominy, B. W.; Chirgwin, J. M. Computer-Graphics Modeling of Human Renin - Specificity, Catalytic Activity and Intron-Exon Junctions. *Febs Letters* **1984**, *174*, 102-111.

(197) Bras, N. F.; Fernandes, P. A.; Ramos, M. J. Molecular dynamics studies on both bound and unbound renin protease. *J Biomol Struct Dyn* **2014**, *32*, 351-363.

(198) Barman, A.; Prabhakar, R. Elucidating the catalytic mechanism of beta-secretase (BACE1): a quantum mechanics/molecular mechanics (QM/MM) approach. *Journal of molecular graphics & modelling* **2013**, *40*, 1-9.

(199) Cascella, M.; Micheletti, C.; Rothlisberger, U.; Carloni, P. Evolutionarily conserved functional mechanics across pepsin-like and retroviral aspartic proteases. *J Am Chem Soc* **2005**, *127*, 3734-3742.

(200) Garrec, J.; Sautet, P.; Fleurat-Lessard, P. Understanding the HIV-1 Protease Reactivity with DFT: What Do We Gain from Recent Functionals? *J Phys Chem B* **2011**, *115*, 8545-8558.

(201) Hong, L.; Tang, J. Flap position of free memapsin 2 (beta-secretase), a model for flap opening in aspartic protease catalysis. *Biochemistry-Us* **2004**, *43*, 4689-4695.

(202) Inoue, I.; Rohrwasser, A.; Helin, C.; Jeunemaitre, X.; Crain, P.; Bohlender, J.; Lifton, R. P.; Corvol, P.; Ward, K.; Lalouel, J. M. A mutation of angiotensinogen in a patient with preeclampsia leads to altered kinetics of the renin-angiotensin system. *J Biol Chem* **1995**, *270*, 11430-11436.

(203) Zhou, A.; Carrell, R. W.; Murphy, M. P.; Wei, Z.; Yan, Y.; Stanley, P. L.; Stein, P. E.; Broughton Pipkin, F.; Read, R. J. A redox switch in angiotensinogen modulates angiotensin release. *Nature* **2010**, *468*, 108-111.

(204) D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R.; Luo, K. M. M., D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S.; Hayik, A. R., G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H.; Gohlke, L. Y., C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C.; Schafmeister, W. S. R., and P.A. Kollman. AMBER9. *University of California* **2006**.

(205) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J Chem Theory Comput* **2006**, *2*, 815-826.

(206) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **1993**, *98*, 5648-5652.

(207) Kohn, W.; Becke, A. D.; Parr, R. G. Density functional theory of electronic structure. *The Journal of Physical Chemistry* **1996**, *100*, 12974-12980.

(208) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J Phys Chem-Us* **1993**, *97*, 10269-10280.

(209) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J Comput Chem* **1990**, *11*, 431-439.

(210) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J Comput Chem* **1984**, *5*, 129-145.

(211) Brás, N. F.; Fernandes, P. A.; Ramos, M. J. QM/MM studies on the β-galactosidase catalytic mechanism: Hydrolysis and transglycosylation reactions. *Journal of Chemical Theory and Computation* **2010**, *6*, 421-433.

(212) Oliveira, E. F.; Cerqueira, N. M.; Ramos, M. J.; Fernandes, P. A. QM/MM study of the mechanism of reduction of 3-hydroxy-3-methylglutaryl coenzyme A catalyzed by human HMG-CoA reductase. *Catalysis Science & Technology* **2016**, *6*, 7172-7185.

(213) Carnevale, V.; Raugei, S.; Piana, S.; Carloni, P. On the nature of the reaction intermediate in the HIV-1 protease: a quantum chemical study. *Computer Physics Communications* **2008**, *179*, 120-123.

(214) Pilote, L.; McKercher, G.; Thibeault, D.; Lamarre, D. Enzymatic characterization of purified recombinant human renin. *Biochemistry and cell biology* **1995**, *73*, 163-170.

(215) Barman, A.; Schurer, S.; Prabhakar, R. Computational modeling of substrate specificity and catalysis of the beta-secretase (BACE1) enzyme. *Biochemistry-Us* **2011**, *50*, 4337-4349.

(216) Bras, N. F.; Ramos, M. J.; Fernandes, P. A. The catalytic mechanism of mouse renin studied with QM/MM calculations. *Phys Chem Chem Phys* **2012**, *14*, 12605-12613.

(217) Hobart, P. M.; Fogliano, M.; O'Connor, B. A.; Schaefer, I. M.; Chirgwin, J. M. Human renin gene: structure and sequence analysis. *Proceedings of the National Academy of Sciences* **1984**, *81*, 5026-5030.

(218) Villar, E. A.; Beglov, D.; Chennamadhavuni, S.; Porco Jr, J. A.; Kozakov, D.; Vajda, S.; Whitty, A. How proteins bind macrocycles. *Nature chemical biology* **2014**, *10*, 723.

(219) Over, B. r.; McCarren, P.; Artursson, P.; Foley, M.; Giordanetto, F.; Grönberg, G.; Hilgendorf, C.; Lee IV, M. D.; Matsson, P. r.; Muncipinto, G. Impact of stereospecific intramolecular hydrogen bonding on cell permeability and physicochemical properties. *Journal of medicinal chemistry* **2014**, *57*, 2746-2754.

(220) Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. The exploration of macrocycles for drug discovery—an underexploited structural class. *Nature Reviews Drug Discovery* **2008**, *7*, 608.

(221) White, C. J.; Yudin, A. K. Contemporary strategies for peptide macrocyclization. *Nature chemistry* **2011**, *3*, 509.

(222) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural product reports* **2013**, *30*, 108-160.

(223) Wu, Z.; Guo, X.; Guo, Z. Sortase A-catalyzed peptide cyclization for the synthesis of macrocyclic peptides and glycopeptides. *Chemical Communications* **2011**, *47*, 9218-9220.

(224) Nguyen, G. K.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J. P. Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nature chemical biology* **2014**, *10*, 732.

(225) Luo, H.; Hong, S.-Y.; Sgambelluri, R. M.; Angelos, E.; Li, X.; Walton, J. D. Peptide macrocyclization catalyzed by a prolyl oligopeptidase involved in α-amanitin biosynthesis. *Chemistry & biology* **2014**, *21*, 1610-1617.

(226) Koehnke, J.; Bent, A.; Houssen, W. E.; Zollman, D.; Morawitz, F.; Shirran, S.; Vendome, J.; Nneoyiegbe, A. F.; Trembleau, L.; Botting, C. H. The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. *Nature Structural and Molecular Biology* **2012**, *19*, 767.

(227) McIntosh, J. A.; Robertson, C. R.; Agarwal, V.; Nair, S. K.; Bulaj, G. W.; Schmidt, E. W. Circular logic: nonribosomal peptide-like macrocyclization with a ribosomal peptide catalyst. *Journal of the American Chemical Society* **2010**, *132*, 15499-15501.

(228) Houssen, W. E.; Jaspars, M. Azole-Based Cyclic Peptides from the Sea Squirt Lissoclinum Patella: Old Scaffolds, New Avenues. *ChemBioChem* **2010**, *11*, 1803-1815.

(229) Šali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology* **1993**, *234*, 779-815.

(230) Dennington, R.; Keith, T.; Millam, J. GaussView, version 5. **2009**.

(231) Alberto, M. E.; Marino, T.; Ramos, M. J.; Russo, N. Atomistic details of the Catalytic Mechanism of Fe (III)− Zn (II) Purple Acid Phosphatase. *Journal of chemical theory and computation* **2010**, *6*, 2424-2433.

(232) Ion, B. F.; Bushnell, E. A.; Luna, P. D.; Gauld, J. W. A molecular dynamics (MD) and quantum mechanics/molecular mechanics (QM/MM) study on ornithine cyclodeaminase (OCD): a tale of two iminiums. *International journal of molecular sciences* **2012**, *13*, 12994-13011.

(233) Lonsdale, R.; J Mulholland, A. QM/MM modelling of drug-metabolizing enzymes. *Current topics in medicinal chemistry* **2014**, *14*, 1339-1347.

(234) Calixto, A. R.; Bras, N. F.; Fernandes, P. A.; Ramos, M. J. Reaction Mechanism of Human Renin Studied by Quantum Mechanics/Molecular Mechanics (QM/MM) Calculations. *Acs Catal* **2014**, *4*, 3869-3876.

(235) Pinto, G. P.; Bras, N. F.; Perez, M. A. S.; Fernandes, P. A.; Russo, N.; Ramos, M. J.; Toscano, M. Establishing the Catalytic Mechanism of Human Pancreatic alpha-Amylase with QM/MM Methods. *Journal of Chemical Theory and Computation* **2015**, *11*, 2508-2516.

(236) Zhang, Y.; Xu, X.; Goddard, W. A. Doubly hybrid density functional for accurate descriptions of nonbond interactions, thermochemistry, and thermochemical kinetics. *Proceedings of the National Academy of Sciences* **2009**, *106*, 4963-4968.

(237) Zhao, Y.; Truhlar, D. G. Density functionals with broad applicability in chemistry. *Accounts Chem Res* **2008**, *41*, 157-167.

(238) Behrendt, L.; Larkum, A. W.; Norman, A.; Qvortrup, K.; Chen, M.; Ralph, P.; Sørensen, S. J.; Trampe, E.; Kühl, M. Endolithic chlorophyll d-containing phototrophs. *The ISME journal* **2011**, *5*, 1072.

(239) Wei, D.; Huang, X.; Liu, J.; Tang, M.; Zhan, C.-G. Reaction pathway and free energy profile for papain-catalyzed hydrolysis of N-acetyl-Phe-Gly 4-nitroanilide. *Biochemistry-Us* **2013**, *52*, 5145-5154.

(240) Gerlt, J. A.; Gassman, P. G. Understanding the rates of certain enzyme-catalyzed reactions: Proton abstraction from carbon acids, acyl transfer reactions, and displacement reactions of phosphodiesters. *Biochemistry-Us* **1993**, *32*, 11943-11952.

(241) Hedstrom, L. Serine protease mechanism and specificity. *Chemical reviews* **2002**, *102*, 4501-4524.

(242) Kraut, J. Serine proteases: structure and mechanism of catalysis. *Annual review of biochemistry* **1977**, *46*, 331-358.

(243) Lee, J.; McIntosh, J.; Hathaway, B. J.; Schmidt, E. W. Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. *Journal of the American Chemical Society* **2009**, *131*, 2122-2124.

(244) Xu, W.; Li, L.; Du, L.; Tan, N. Various mechanisms in cyclopeptide production from precursors synthesized independently of non-ribosomal peptide synthetases. *Acta Biochim Biophys Sin* **2011**, *43*, 757-762.

(245) Kohen, A.; Klinman, J. P. Enzyme catalysis: Beyond classical paradigms. *Accounts Chem Res* **1998**, *31*, 397-404.

(246) Bruice, T. C. A view at the millennium: The efficiency of enzymatic catalysis. *Accounts Chem Res* **2002**, *35*, 139-148.

(247) Gao, J. L.; Ma, S. H.; Major, D. T.; Nam, K.; Pu, J. Z.; Truhlar, D. G. Mechanisms and free energies of enzymatic reactions. *Chemical Reviews* **2006**, *106*, 3188-3209.

(248) Mulholland, A. J. Modelling enzyme reaction mechanisms, specificity and catalysis. *Drug Discov Today* **2005**, *10*, 1393-1402.

(249) Ranaghan, K. E.; Mulholland, A. J. Investigations of enzyme-catalysed reactions with combined quantum mechanics/molecular mechanics (QM/MM) methods. *Int Rev Phys Chem* **2010**, *29*, 65-133.

(250) Swiderek, K.; Tunon, I.; Moliner, V. Predicting enzymatic reactivity: from theory to design. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014**, *4*, 407-421.

(251) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. A practical guide to modelling enzyme-catalysed reactions. *Chem Soc Rev* **2012**, *41*, 3025-3038.

(252) Gherib, R.; Dokainish, H. M.; Gauld, J. W. Multi-Scale Computational Enzymology: Enhancing Our Understanding of Enzymatic Catalysis. *International Journal of Molecular Sciences* **2014**, *15*, 401-422.

(253) Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. The ONIOM method: its foundation and applications to metalloenzymes and photobiology. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 327-350.

(254) Siegbahn, P. E.; Himo, F. Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *JBIC Journal of Biological Inorganic Chemistry* **2009**, *14*, 643-651.

(255) Leopoldini, M.; Marino, T.; Michelini, M. D.; Rivalta, I.; Russo, N.; Sicilia, E.; Toscano, M. The role of quantum chemistry in the elucidation of the elementary mechanisms of catalytic processes: from atoms, to surfaces, to enzymes. *Theor Chem Acc* **2007**, *117*, 765-779.

(256) Shurki, A.; Derat, E.; Barrozo, A.; Kamerlin, S. C. L. How valence bond theory can help you understand your (bio) chemical reaction. *Chem Soc Rev* **2015**, *44*, 1037-1052.

(257) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc Natl Acad Sci U S A* **2002**, *99*, 12562-12566.

(258) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z. F.; Liu, F. Y.; Li, H. B.; Ding, L. N.; Morokuma, K. The ONIOM Method and Its Applications. *Chemical Reviews* **2015**, *115*, 5678-5796.

(259) Neves, R. P. P.; Fernandes, P. A.; Ramos, M. J. Unveiling the Catalytic Mechanism of NADP(+)-Dependent Isocitrate Dehydrogenase with QM/MM Calculations. *Acs Catal* **2016**, *6*, 357-368.

(260) Moreira, C.; Ramos, M. J.; Fernandes, P. A. Reaction Mechanism of Mycobacterium Tuberculosis Glutamine Synthetase Using Quantum Mechanics/Molecular Mechanics Calculations. *Chem-Eur J* **2016**, *22*, 9218-9225.

(261) Medina, F. E.; Neves, R. P.; Ramos, M. J.; Fernandes, P. A. A QM/MM study of the reaction mechanism of human beta-ketoacyl reductase. *Phys Chem Chem Phys* **2016**, *19*, 347-355.

(262) Hamada, Y.; Kanematsu, Y.; Tachikawa, M. QM/MM Study on Sialyltransferase Reaction Mechanism. *Biochemistry-Us* **2016**.

(263) Reis, M.; Alves, C. N.; Lameira, J.; Tunon, I.; Marti, S.; Moliner, V. The catalytic mechanism of glyceraldehyde 3-phosphate dehydrogenase from Trypanosoma cruzi elucidated via the QM/MM approach. *Phys Chem Chem Phys* **2013**, *15*, 3772-3785.

(264) Perez-Gallegos, A.; Garcia-Viloca, M.; Gonzalez-Lafont, A.; Lluch, J. M. A QM/MM study of the associative mechanism for the phosphorylation reaction catalyzed by protein kinase A and its D166A mutant. *J Comput Aid Mol Des* **2014**, *28*, 1077-1091.

(265) Bravaya, K. B.; Subach, O. M.; Korovina, N.; Verkhusha, V. V.; Krylov, A. I. Insight into the Common Mechanism of the Chromophore Formation in the Red Fluorescent Proteins: The Elusive Blue Intermediate Revealed. *J Am Chem Soc* **2012**, *134*, 2807-2814.

(266) Mlynsky, V.; Walter, N. G.; Sponer, J.; Otyepka, M.; Banas, P. The role of an active site Mg2+ in HDV ribozyme self-cleavage: insights from QM/MM calculations. *Phys Chem Chem Phys* **2015**, *17*, 670-679.

(267) Lodola, A.; Mor, M.; Hermann, J. C.; Tarzia, G.; Piomelli, D.; Mulholland, A. J. QM/MM modelling of oleamide hydrolysis in fatty acid amide hydrolase (FAAH) reveals a new mechanism of nucleophile activation. *Chemical Communications* **2005**, 4399-4401.

(268) Zhang, S. J.; Ma, G. C.; Liu, Y. J.; Ling, B. P. Theoretical study of the hydrolysis mechanism of 2-pyrone-4,6-dicarboxylate (PDC) catalyzed by LigI. *Journal of molecular graphics & modelling* **2015**, *61*, 21-29.

(269) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J Mol Struc-Theochem* **1999**, *461*, 1-21.

(270) Jeffrey, C. J.; Gloss, L. M.; Petsko, G. A.; Ringe, D. The role of residues outside the active site: structural basis for function of C191 mutants of Escherichia coli aspartate aminotransferase. *Protein Eng* **2000**, *13*, 105-112.

(271) Shen, Y. L.; Li, X.; Chai, T. Y.; Wang, H. Outer-sphere residues influence the catalytic activity of a chalcone synthase from Polygonum cuspidatum. *Febs Open Bio* **2016**, *6*, 610-618.

(272) Ozer, N.; Schiffer, C. A.; Haliloglu, T. Rationale for more diverse inhibitors in competition with substrates in HIV-1 protease. *Biophys J* **2010**, *99*, 1650-1659.

(273) Yang, L.; Song, G.; Carriquiry, A.; Jernigan, R. L. Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure* **2008**, *16*, 321-330.

(274) Davies, D. R. The Structure and Function of the Aspartic Proteinases. *Annu Rev Biophys Bio* **1990**, *19*, 189-215.

(275) Fitzgerald, P. M. D.; Springer, J. P. Structure and Function of Retroviral Proteases. *Annu Rev Biophys Bio* **1991**, *20*, 299-320.

(276) Krzeminska, A.; Moliner, V.; Swiderek, K. Dynamic and Electrostatic Effects on the Reaction Catalyzed by HIV-1 Protease. *J Am Chem Soc* **2016**, *138*, 16283-16298.

(277) Chatfield, D. C.; Eurenius, K. P.; Brooks, B. R. HIV-1 protease cleavage mechanism: A theoretical investigation based on classical MD simulation and reaction path calculations using a hybrid QM/MM potential. *Theochem-J Mol Struc* **1998**, *423*, 79-92.

(278) Trylska, J.; Bala, P.; Geller, M.; Grochowski, P. Molecular dynamics simulations of the first steps of the reaction catalyzed by HIV-1 protease. *Biophysical Journal* **2002**, *83*, 794-807.

(279) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B. H.; Wlodawer, A. Structure of Complex of Synthetic Hiv-1 Protease with a Substrate-Based Inhibitor at 2.3-a Resolution. *Science* **1989**, *246*, 1149-1152.

(280) Northrop, D. B. Follow the protons: A low-barrier hydrogen bond unifies the mechanisms of the aspartic proteases. *Accounts Chem Res* **2001**, *34*, 790-797.

(281) Polgar, L.; Szeltner, Z.; Boros, I. Substrate-Dependent Mechanisms in the Catalysis of Human-Immunodeficiency-Virus Protease. *Biochemistry-Us* **1994**, *33*, 9351-9357.

(282) Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes. *Structure* **2002**, *10*, 369-381.

(283) Rodriguez, E. J.; Angeles, T. S.; Meek, T. D. Use of nitrogen-15 kinetic isotope effects to elucidate details of the chemical mechanism of human immunodeficiency virus 1 protease. *Biochemistry-Us* **1993**, *32*, 12380-12385.

(284) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *Journal of Chemical Physics* **1995**, *103*, 8577-8593.

(285) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, *3*, 198-210.

(286) Hur, S.; Bruice, T. C. Comparison of formation of reactive conformers (NACs) for the Claisen rearrangement of chorismate to prephenate in water and in the E. coli mutase: the efficiency of the enzyme catalysis. *J Am Chem Soc* **2003**, *125*, 5964-5972.

(287) Guo, H.; Cui, Q.; Lipscomb, W. N.; Karplus, M. Substrate conformational transitions in the active site of chorismate mutase: their role in the catalytic mechanism. *Proc Natl Acad Sci U S A* **2001**, *98*, 9032-9037.

(288) Roca, M.; De Maria, L.; Wodak, S. J.; Moliner, V.; Tunon, I.; Giraldo, J. Coupling of the guanosine glycosidic bond conformation and the ribonucleotide cleavage reaction: implications for barnase catalysis. *Proteins* **2008**, *70*, 415-428.

(289) Ruscio, J. Z.; Kohn, J. E.; Ball, K. A.; Head-Gordon, T. The influence of protein dynamics on the success of computational enzyme design. *J Am Chem Soc* **2009**, *131*, 14111-14115.

(290) van der Kamp, M. W.; Sirirak, J.; Zurek, J.; Allemann, R. K.; Mulholland, A. J. Conformational change and ligand binding in the aristolochene synthase catalytic cycle. *Biochemistry-Us* **2013**, *52*, 8094-8105.

(291) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. Compound I Reactivity Defines Alkene Oxidation Selectivity in Cytochrome P450cam. *J Phys Chem B* **2010**, *114*, 1156-1162.

(292) Lodola, A.; Sirirak, J.; Fey, N.; Rivara, S.; Mor, M.; Mulholland, A. J. Structural Fluctuations in Enzyme-Catalyzed Reactions: Determinants of Reactivity in Fatty Acid Amide Hydrolase from Multivariate Statistical Analysis of Quantum Mechanics/Molecular Mechanics Paths. *J Chem Theory Comput* **2010**, *6*, 2948-2960.

(293) Zhang, Y. K.; Kua, J.; McCammon, J. A. Influence of structural fluctuation on enzyme reaction energy barriers in combined quantum mechanical/molecular mechanical studies. *J Phys Chem B* **2003**, *107*, 4459-4463.

(294) Hu, P.; Zhang, Y. K. Catalytic mechanism and product specificity of the histone lysine methyltransferase SET7/9: An ab initio QM/MM-FE study with multiple initial structures. *Journal of the American Chemical Society* **2006**, *128*, 1272-1278.

(295) Benkovic, S. J.; Hammes, G. G.; Hammes-Schiffer, S. Free-energy landscape of enzyme catalysis. *Biochemistry-Us* **2008**, *47*, 3317-3321.

(296) Ribeiro, A. J. M.; Ramos, M. J.; Fernandes, P. A. The Catalytic Mechanism of HIV-1 Integrase for DNA 3 '-End Processing Established by QM/MM Calculations. *Journal of the American Chemical Society* **2012**, *134*, 13436-13447.

(297) Sanchez-Martinez, M.; Marcos, E.; Tauler, R.; Field, M.; Crehuet, R. Conformational Compression and Barrier Height Heterogeneity in the N-Acetylglutamate Kinase. *J Phys Chem B* **2013**, *117*, 14261-14272.

(298) Sousa, R. P.; Fernandes, P. A.; Ramos, M. J.; Bras, N. F. Insights into the reaction mechanism of 3-O-sulfotransferase through QM/MM calculations. *Phys Chem Chem Phys* **2016**, *18*, 11488-11496.

(299) Li, Y. W.; Zhang, R. M.; Du, L. K.; Zhang, Q. Z.; Wang, W. X. How Many Conformations of Enzymes Should Be Sampled for DFT/MM Calculations? A Case Study of Fluoroacetate Dehalogenase. *International Journal of Molecular Sciences* **2016**, *17*.

(300) Loerbroks, C.; Heimermann, A.; Thiel, W. Solvents effects on the mechanism of cellulose hydrolysis: A QM/MM study. *J Comput Chem* **2015**, *36*, 1114-1123.

(301) Cooper, A. M.; Kastner, J. Averaging Techniques for Reaction Barriers in QM/MM Simulations. *Chemphyschem* **2014**, *15*, 3264-3269.

(302) Piana, S.; Carloni, P.; Parrinello, M. Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease. *J Mol Biol* **2002**, *319*, 567-583.

(303) Turner, A. J.; Moliner, V.; Williams, I. H. Transition-state structural refinement with GRACE and CHARMM: Flexible QM/MM modelling for lactate dehydrogenase. *Phys Chem Chem Phys* **1999**, *1*, 1323-1331.

(304) Palermo, G.; Campomanes, P.; Neri, M.; Piomelli, D.; Cavalli, A.; Rothlisberger, U.; De Vivo, M. Wagging the tail: essential role of substrate flexibility in FAAH catalysis. *Journal of chemical theory and computation* **2013**, *9*, 1202-1213.

(305) Genna, V.; Gaspari, R.; Dal Peraro, M.; De Vivo, M. Cooperative motion of a key positively charged residue and metal ions for DNA replication catalyzed by human DNA Polymerase-η. *Nucleic Acids Res* **2016**, *44*, 2827-2836.

(306) Dan, N. Understanding dynamic disorder fluctuations in single-molecule enzymatic reactions. *Curr Opin Colloid In* **2007**, *12*, 314-321.

(307) Terentyeva, T. G.; Engelkamp, H.; Rowan, A. E.; Komatsuzaki, T.; Hofkens, J.; Li, C. B.; Blank, K. Dynamic Disorder in Single-Enzyme Experiments: Facts and Artifacts. *Acs Nano* **2012**, *6*, 346-354.

(308) Ruiz-Pernia, J. J.; Luk, L. Y. P.; Garcia-Meseguer, R.; Marti, S.; Loveridge, E. J.; Tunon, I.; Moliner, V.; Allemann, R. K. Increased Dynamic Effects in a Catalytically Compromised Variant of Escherichia coli Dihydrofolate Reductase. *Journal of the American Chemical Society* **2013**, *135*, 18689-18696.

(309) Bras, N. F.; Cerqueira, N. M. F. S. A.; Ramos, M. J.; Fernandes, P. A. Glycosidase inhibitors: a patent review (2008-2013). *Expert Opin Ther Pat* **2014**, *24*, 857-874.

(310) Brayer, G. D.; Sidhu, G.; Maurus, R.; Rydberg, E. H.; Braun, C.; Wang, Y. L.; Nguyen, N. T.; Overall, C. H.; Withers, S. G. Subsite mapping of the human pancreatic alpha-amylase active site through structural, kinetic, and mutagenesis techniques. *Biochemistry-Us* **2000**, *39*, 4778-4791.

(311) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712-725.

(312) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A generalizable Biomolecular force field. Carbohydrates. *J Comput Chem* **2008**, *29*, 622-655.

(313) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics & modelling* **2006**, *25*, 247-260.

(314) Darden, T.; Perera, L.; Li, L.; Pedersen, L. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* **1999**, *7*, R55-60.

(315) Crowley, M. F.; Darden, T. A.; Cheatham, T. E.; Deerfield, D. W. Adventures in improving the scaling and accuracy of a parallel molecular dynamics program. *J Supercomput* **1997**, *11*, 255-278.

(316) Raghavachari, K. Perspective on "Density functional thermochemistry. III. The role of exact exchange" - Becke AD (1993) J Chem Phys 98:5648-52. *Theor Chem Acc* **2000**, *103*, 361-363.

(317) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **2013**, *9*, 3084-3095.

(318) Calixto, A. R.; Ramos, M. J.; Fernandes, P. A. Influence of Frozen Residues on the Exploration of the PES of Enzyme Reaction Mechanisms. *J Chem Theory Comput* **2017**, *13*, 5486-5495.

(319) Pereira, A. T.; Ribeiro, A. J.; Fernandes, P. A.; Ramos, M. J. Benchmarking of density functionals for the kinetics and thermodynamics of the hydrolysis of glycosidic bonds catalyzed by glycosidases. *International Journal of Quantum Chemistry* **2017**.

(320) Vasilevskaya, T.; Khrenova, M. G.; Nemukhin, A. V.; Thiel, W. Methodological aspects of QM/MM calculations: A case study on matrix metalloproteinase-2. *J Comput Chem* **2016**, *37*, 1801-1809.

(321) Li, Y.; Zhang, R.; Du, L.; Zhang, Q.; Wang, W. Catalytic mechanism of C–F bond cleavage: insights from QM/MM analysis of fluoroacetate dehalogenase. *Catalysis Science & Technology* **2016**, *6*, 73-80.

(322) van der Kamp, M. W.; Zurek, J.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. Testing high-level QM/MM methods for modeling enzyme reactions: acetyl-CoA deprotonation in citrate synthase. *J Phys Chem B* **2010**, *114*, 11303-11314.

(323) Palmer, A. G., 3rd. Enzyme dynamics from NMR spectroscopy. *Acc Chem Res* **2015**, *48*, 457-465.

(324) Lu, H. P.; Xun, L.; Xie, X. S. Single-molecule enzymatic dynamics. *Science* **1998**, *282*, 1877-1882.

(325) Antikainen, N. M.; Smiley, R. D.; Benkovic, S. J.; Hammes, G. G. Conformation coupled enzyme catalysis: single-molecule and transient kinetics investigation of dihydrofolate reductase. *Biochemistry-Us* **2005**, *44*, 16835-16843.

(326) Smiley, R. D.; Hammes, G. G. Single molecule studies of enzyme mechanisms. *Chemical reviews* **2006**, *106*, 3080-3094.

(327) Nashine, V. C.; Hammes-Schiffer, S.; Benkovic, S. J. Coupled motions in enzyme catalysis. *Current opinion in chemical biology* **2010**, *14*, 644-651.

(328) Santos-Martins, D.; Calixto, A. R.; Fernandes, P. A.; Ramos, M. J. A Buried Water Molecule Influences Reactivity in α-Amylase on a Subnanosecond Time Scale. *Acs Catal* **2018**, *8*, 4055-4063.

(329) Mata, R. A.; Werner, H. J.; Thiel, S.; Thiel, W. Toward accurate barriers for enzymatic reactions: QM/MM case study on p-hydroxybenzoate hydroxylase. *J Chem Phys* **2008**, *128*, 025104.

(330) Lonsdale, R.; Hoyle, S.; Grey, D. T.; Ridder, L.; Mulholland, A. J. Determinants of reactivity and selectivity in soluble epoxide hydrolase from quantum mechanics/molecular mechanics modeling. *Biochemistry-Us* **2012**, *51*, 1774-1786.

(331) Sokkar, P.; Boulanger, E.; Thiel, W.; Sanchez-Garcia, E. Hybrid quantum mechanics/molecular mechanics/coarse grained modeling: A triple-resolution approach for biomolecular systems. *Journal of chemical theory and computation* **2015**, *11*, 1809-1818.

(332) Li, Y.; Shi, X.; Zhang, Q.; Hu, J.; Chen, J.; Wang, W. Computational evidence for the detoxifying mechanism of epsilon class glutathione transferase toward the insecticide DDT. *Environmental science & technology* **2014**, *48*, 5008-5016.

(333) Abad, E.; Zenn, R. K.; Kästner, J. Reaction mechanism of monoamine oxidase from QM/MM calculations. *The Journal of Physical Chemistry B* **2013**, *117*, 14238-14246.

(334) Christov, C. Z.; Lodola, A.; Karabencheva-Christova, T. G.; Wan, S.; Coveney, P. V.; Mulholland, A. J. Conformational effects on the pro-S hydrogen abstraction reaction in cyclooxygenase-1: an integrated QM/MM and MD study. *Biophysical journal* **2013**, *104*, L5-L7.

(335) Hu, L.; Söderhjelm, P. r.; Ryde, U. Accurate reaction energies in proteins obtained by combining QM/MM and large QM calculations. *Journal of chemical theory and computation* **2012**, *9*, 640-649.

(336) Schoneboom, J. C.; Lin, H.; Reuter, N.; Thiel, W.; Cohen, S.; Ogliaro, F.; Shaik, S. The elusive oxidant species of cytochrome P450 enzymes: characterization by combined quantum mechanical/molecular mechanical (QM/MM) calculations. *J Am Chem Soc* **2002**, *124*, 8142-8151.

(337) Rosenthal, R. G.; Vögeli, B.; Wagner, T.; Shima, S.; Erb, T. J. A conserved threonine prevents self-intoxication of enoyl-thioester reductases. *Nature chemical biology* **2017**, *13*, 745.

(338) Ryde, U. How many conformations need to be sampled to obtain converged QM/MM energies? The curse of exponential averaging. *Journal of chemical theory and computation* **2017**, *13*, 5745-5752.

(339) Wong, K. Y.; Gao, J. The reaction mechanism of paraoxon hydrolysis by phosphotriesterase from combined QM/MM simulations. *Biochemistry-Us* **2007**, *46*, 13352-13369.

(340) Repic, M.; Vianello, R.; Purg, M.; Duarte, F.; Bauer, P.; Kamerlin, S. C.; Mavri, J. Empirical valence bond simulations of the hydride transfer step in the monoamine oxidase B catalyzed metabolism of dopamine. *Proteins* **2014**, *82*, 3347-3355.

(341) Maršavelski, A.; Petrović, D. a.; Bauer, P.; Vianello, R.; Kamerlin, S. C. L. Empirical Valence Bond Simulations Suggest a Direct Hydride Transfer Mechanism for Human Diamine Oxidase. *ACS Omega* **2018**, *3*, 3665-3674.

(342) Barrozo, A.; Liao, Q.; Esguerra, M.; Marloie, G.; Florian, J.; Williams, N. H.; Kamerlin, S. C. L. Computer simulations of the catalytic mechanism of wild-type and mutant beta-phosphoglucomutase. *Org Biomol Chem* **2018**, *16*, 2060-2073.

(343) Hu, P.; Zhang, Y. Catalytic mechanism and product specificity of the histone lysine methyltransferase SET7/9: an ab initio QM/MM-FE study with multiple initial structures. *J Am Chem Soc* **2006**, *128*, 1272-1278.

(344) Piana, S.; Carloni, P. Conformational flexibility of the catalytic Asp dyad in HIV-1 protease: An ab initio study on the free enzyme. *Proteins: Structure, Function, and Bioinformatics* **2000**, *39*, 26-36.

(345) Brás, N. F.; Ramos, M. J.; Fernandes, P. A. The catalytic mechanism of mouse renin studied with QM/MM calculations. *Phys Chem Chem Phys* **2012**, *14*, 12605-12613.