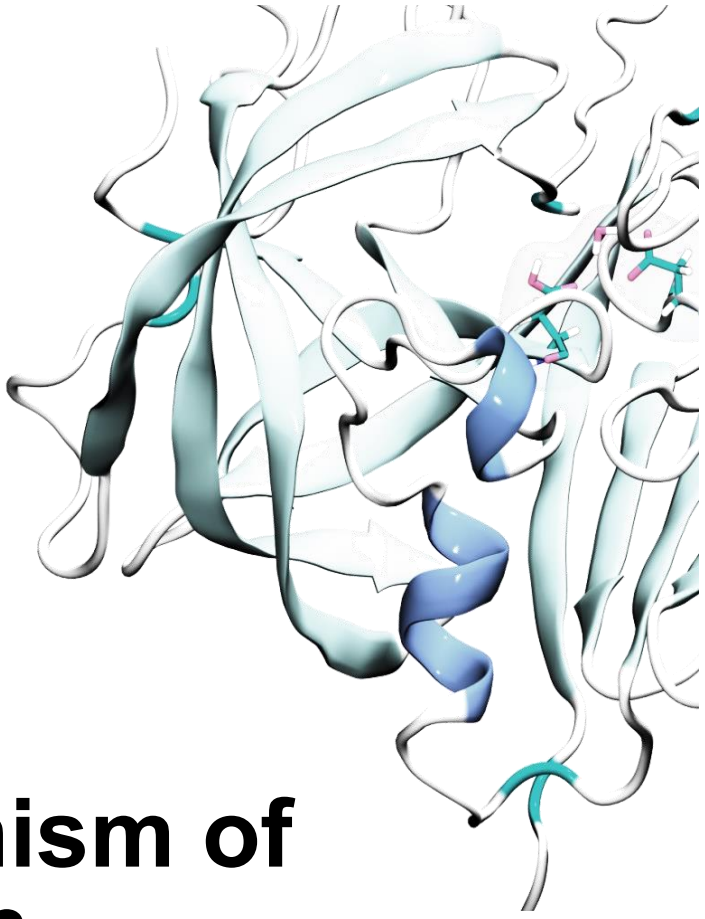


Elucidating the catalytic mechanism of human renin with hybrid Quantum Mechanics and Molecular Mechanics studies

Ana Rita Almeida Calixto Silva

Mestrado em Bioquímica
Departamento de Química e Bioquímica
2013

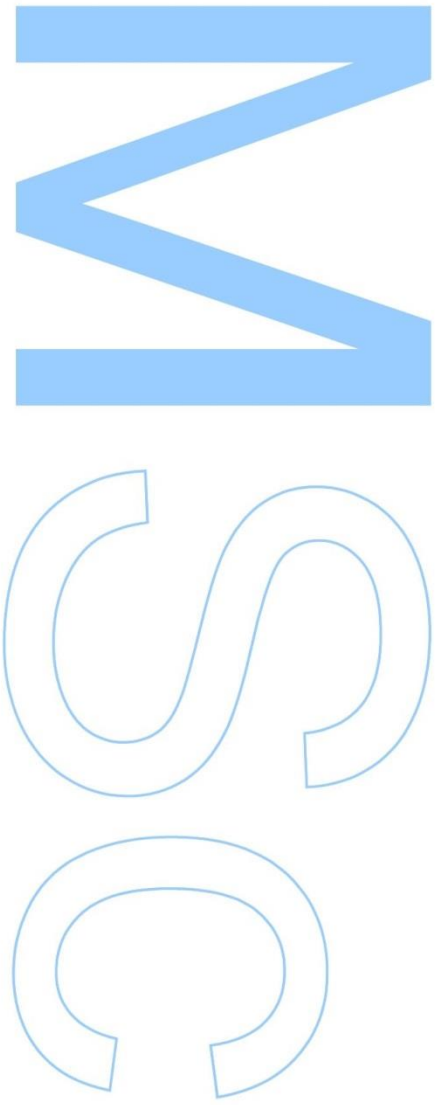
Orientadora
Doutora Maria João Ramos, Professora Catedrática,
Faculdade de Ciências da Universidade do Porto

Coorientadora
Doutora Natércia Brás, Investigadora
Faculdade de Ciências da Universidade do Porto

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

U.PORTO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

Todas as correções determinadas
pelo júri, e só essas, foram efetuadas.
O Presidente do Júri,
Porto, _____/_____/_____



The work presented here has contributed to the following conference presentations:

Oral Communication

Ana Rita Calixto, Natércia Fernandes Brás, Pedro Alexandrino Fernandes and Maria João Ramos. *"The catalytic mechanism of human renin protease – a quantum mechanics /molecular mechanics study"*, IJUP 2013 – 6º Encontro de Investigação Jovem da Universidade do Porto, 13th-15th February 2013.

Poster presentation

Ana Rita Calixto, Natércia Fernades Brás, Pedro Alexandrino Fernandes and Maria João Ramos, *"Elucidating the catalytic mechanism of human renin with a QM/MM study"* XXIII Encontro Nacional da SPQ - Desafios em Química, Universidade de Aveiro, 12th-14th June 2013

Ana Rita Calixto, Natércia Fernades Brás, Pedro Alexandrino Fernandes and Maria João Ramos, *"A QM/MM investigation of the catalytic mechanism of human renin protease"* , XVIII Encontro Luso Galego de Química, Universidade deTrás-os-Montes e Alto Douro, Vila Real, 28th-30th November 2012

Ana Rita Calixto, Natércia Fernades Brás, Pedro Alexandrino Fernandes and Maria João Ramos, *"Computational study of the catalytic mechanism of human renin protease"*, 5º Encontro de Investigação Jovem da Universidade do Porto, 22nd-24th February 2012.

Acknowledgments

"*Appreciation is a wonderful thing: It makes what is excellent in others belong to us as well.*"

Voltaire

First of all, I would like to express my sincere gratitude to my supervisor, Professor Maria João Ramos, for giving me the opportunity to work with her, for her excellent guidance, continuous support, patience and constant motivation. Thank you for all the opportunities and suggestions during this work and for sharing your wealth knowledge with me.

I am also heartily thankful to Professor Pedro Alexandrino Fernandes, whose guidance, enthusiasm and encouragement to recover when my steps faltered were essential to finish my dissertation. His suggestions were always an asset to progress of this work. Thank you for have introduced me to the fascinating world of Computational Biochemistry and for the opportunity to work in your group.

I would also like to thank Natércia, my co-supervisor, for guiding me throughout the whole work and for all precious teachings and advices. I am deeply grateful for your support, patient, understanding and generosity. Thank you very much for your availability for everything and for keeping my hopes up when the results did not seems good.

I extend my appreciation to all Theoretical Chemistry and Computational Biochemistry group for providing me an excellent atmosphere to work. I am very thankful for the way I was received in the group. I would like to thank Rui and João for all the fun times we have had during the last year that distracting me from my work, for all the advices and help. I am also thankful to Rui Sousa, Cátia, Rita, Eduardo, Sílvia, Óscar and Zé, not only for all the support in this work, but also for the excellent work environment that you promote and, of course, for all the fun coffee breaks that increase my productivity every day.

I would also like to thank Silvia and Lau for all their friendship, understanding and support during these years.

I owe my deepest gratitude to Damien, who has always stood by me in the best and worst moments. Thank you for your care, patience and support every time I needed.

I am most grateful to my mother and brother for all the sacrifices during the last years and for all the unconditional love and support they have always give me. Without whom this work would not have been possible.

Finally, I would like to thank Foundation of Science and Technology of Portugal (FCT) for the financial support (project PTDC/QUI-QUI/121744/2010).

Abstract

Hypertension, also referred as high blood pressure, is a chronical condition that affects nearly 25% of adults worldwide. The increasing prevalence of this illness has contributed to the pandemic of cardiovascular and renal diseases, which are responsible for a large number of deaths every year [1]. Therefore, it is fundamental to develop new antihypertensive drugs, in order to control and prevent the complications associated with high blood pressure.

The Renin-Angiotensin System (RAS) is the major regulating system of the blood pressure and it consists of a two-step cascade. Firstly, the aspartic protease renin cleaves its only known substrate, angiotensinogen, in a rate limiting step, leading to the formation of the decapeptide angiotensin I (Ang I). Secondly, Ang I is transformed by angiotensin-converting enzyme (ACE) to produce the octapeptide angiotensin II (Ang II), which binds to angiotensin II receptors (AT receptors) and mediates the blood pressure control. Effective antihypertensive drugs involving different RAS targets have already been developed. However, it is believed that the inhibition of renin would offer better results in blood pressure control, due to its specificity for only one substrate and the fewer adverse treatment side effects [2, 3].

Therefore, the main goal of the present work is to describe, with computational methods, the atomistic details of the catalytic mechanism of human renin, to allow future studies on its inhibition. In order to achieve the purpose of this work the initial system (renin + angiotensinogen) was divided in two layers that were studied at different theoretical levels (Density functional theory (DFT) and Molecular Mechanics (MM)). The geometries were optimized with the ONIOM methodology, at the B3LYP/6-31G(d):Amber) level. The energies of the stationary points were recalculated with single-point calculations, using a large basis set (6-311++G(2d,2p)) and a progressive increase in the number of atoms in the DFT layer. Molecular dynamics (MD) simulations were also performed in order to understand the behavior of the system along the time.

Our results suggest that the angiotensinogen hydrolysis by renin occurs by an acid/base mechanism, through three elementary steps. It begins with the formation of a stable gem-diol intermediate, followed by the scissile bond nitrogen protonation and it ends with the

completely cleavage of the peptide bond. The final reaction products are two peptides with carboxylic acid and amine extremities.

We observed that the formation of the gem-diol intermediate is rate limiting, with a barrier near 20 kcal.mol^{-1}. We also conclude that the residues around the active center greatly influence the catalytic mechanism of this enzyme. In addition, we also confirm that human renin catalytic mechanism is very similar to the mouse one [4].

The data obtained in the present work provides important clues about renin catalytic mechanism, which will enable further studies on its inhibition and, consequently, the development of new antihypertensive drugs.

Key words: Hypertension, Renin Angiotensin System, Renin, Angiotensinogen, Catalytic Mechanism, Computational methods, ONIOM

Resumo[1]

A hipertensão arterial, também conhecida como pressão sanguínea elevada, é uma doença crónica que afeta cerca de 25% da população mundial. A prevalência desta condição patológica tem contribuído para um aumento significativo de doenças cardiovasculares e renais, sendo estas responsáveis por uma grande percentagem de mortes todos os anos [1]. Tendo isto em conta, de forma a controlar esta condição e a prevenir possíveis complicações a ela associadas, torna-se fundamental o desenvolvimento de novos fármacos anti-hipertensivos que apresentem uma melhor resposta, face aos já existentes.

O sistema renina-angiotensina (RAS) é o principal sistema de regulação da pressão sanguínea e consiste numa cascata que pode ser resumida em dois passos. A primeira enzima envolvida nesta cascata é uma protease aspártica, a renina, sendo esta responsável pela clivagem do angiotensinogénio, o seu único substrato conhecido. Este é um passo limitante de toda a cascata e origina a formação do decapeptido Angiotensina I (Ang I). Num segundo passo, a Ang I é clivada pela enzima conversora da angiotensina (ACE) levando à formação do octapeptido angiotensina II (Ang II) que, por sua vez, interage com os recetores de Ang II, mediando, assim o controlo da pressão sanguínea. Existem no mercado vários fármacos que atuam sobre diferentes alvos do RAS, no entanto, acredita-se que o melhor alvo a inibir nesta cascata será a enzima renina, dada a sua especificidade para apenas um substrato e, uma vez que a sua inibição parece estar associada ao aparecimento de menos efeitos adversos [2, 3] .

Tendo isto em conta, o principal objetivo do presente trabalho passa por tentar descrever, através de metodologia computacionais e com detalhe atomístico, o mecanismo catalítico da enzima renina de humano, de forma a permitir e facilitar futuros estudos da sua inibição. Sendo assim, de forma a atingir este objetivo, o sistema em estudo foi divido em duas camadas que, por sua vez, foram estudadas com diferentes níveis teóricos (Teoria do funcional de densidade (DFT) e mecânica molecular (MM). As geometrias foram otimizadas com o método ONIOM com nível teórico B3LYP/6-31G(d):Amber. Nos pontos estacionários,

[1] As abreviaturas usadas neste resumo derivam do termo em inglês

as energias forma recalculadas usando uma base de funções maior (6-311++G(2d,2p)) e incluído, progressivamente, na camada DFT um maior numero de átomos. Foi ainda realizada uma simulação de Dinâmica Molecular, com o intuito de perceber o comportamento do sistema ao longo do tempo.

Os resultados obtidos sugerem que a hidrólise do angiotensinogénio, pela enzima renina, ocorre através de um mecanismo acido-base, baseado em três passos elementares. Num primeiro passo, ocorre a formação de um intermediário gem-diol, seguida por um segundo passo de protonação do azoto da ligação a quebrar. Num terceiro, e último passo, ocorre a clivagem completa da ligação peptídica, sendo os produtos finais da reação dois péptidos com extremidades ácido carboxílico e amina.

Pode ainda ser verificado que a formação do intermediário gem-diol é limitante da velocidade da reação e que está associada a uma barreira energética de cerca de 20 $kcal.mol^{-1}$. Conclui-se ainda que os aminoácidos em torno do centro ativo parecem influenciar, de forma significativa, a energia associada a esta reação. Para além disso, verificou-se, ainda, que o mecanismo pelo qual a enzima humana cliva o seu substrato é semelhante ao seguido pela enzima de ratinho [4].

Os resultados obtidos neste trabalho contém informações essenciais sobre o detalhe do mecanismo da renina, que poderão ser úteis em futuros estudos de inibição desta enzima e, consequentemente, no desenvolvimento de novos fármacos anti-hipertensivos.

Palavras-chave: Hipertensão, Sistema Renina Angiotensina, Renina, Angiotensinogénio, Mecanismo Catalítico, Métodos Computacionais, ONIOM

Table of contents

Acknowledgments iv
Abstract vii
Resumo ix
Abbreviations xviii

CHAPTER 1 - INTRODUCTION 1
1.1 Introductory note 1
1.2 Hypertension 2
1.3 Renin-Angiotensin System 3
1.3.1 Renin, Angiotensinogen and Angiotensin I 3
1.3.2 Angiotensin Converting Enzyme and Angiotensin II 5
1.3.3 Angiotensin receptors 6
1.4 Anti-hypertensive Agents 7
1.4.1 Diuretics 7
1.4.2 β-blockers 8
1.4.3 Calcium channel blockers 8
1.4.4 Vaccination 8
1.4.5 RAS inhibition 8
1.4.5.1 Renin inhibitors 9
1.4.5.2 ACE inhibitors 9
1.4.5.2 AT_1 receptor antagonists 9
1.5 Renin 11
1.5.1 Relation between structure and function 11
1.5.1 Proposal for renin catalytic mechanism 13
1.5.3 Development of renin inhibitors 14
1.6 Computational enzymatic catalysis and drug design 15
1.6.1 The role of computational chemistry in enzymatic catalysis 16
1.6.2 Rational drug design 17
1.7 Purpose of the present work 17

CHAPTER 2 - THEORETICAL BACKGROUND 19
2.1 Introduction 19
2.2 Molecular Mechanics 21
2.2.1 Basic Theory – The force field energy 22
2.2.1.1 The Stretch bond energy 23
2.2.1.2 The bending energy for angles 24
2.2.1.3 The dihedral / torsion energy 24
2.2.1.4 The van der Waals energy 25
2.2.1.3 The electrostatic energy 26
2.2.2 Molecular dynamics 27
2.3 Quantum Mechanics 28
2.3.1 The Schrödinger equation 29
2.3.2 The Hamiltonian 30
2.3.3 The Born-Oppenheimer Approximation 31
2.3.4 Hartree-Fock Theory 32
2.3.5 Linear combination of atomic orbitals 34
2.4 Density Functional Theory 35
2.4.1 The Hohenberg-Kohn Theorem 35
2.5.2 The Konh-Sham Theorem 36
2.5 Basis set 38
2.5.1 Slater type orbitals 38
2.5.2 Gaussian type orbitals 38
2.6 Hybrid QM/MM methods 40
2.6.1 The ONIOM method 41
2.6.1.1 Mechanical embedding scheme 43
2.6.1.2 Electrostatic embedding scheme 43
2.6.1.3 QM/MM boundary and link atoms 44

CHAPTER 3 - COMPUTATIONAL METHODOLOGY 45
3.1 Introductory note 45
3.2 Preparation of the system 46
3.2.1 Ren:$Ang_{dodecapetide}$:W1 model 46

3.2.2 Ren:Ang$_{dodecapetide}$:W$_1$:W$_2$ model 46
3.2.3 Ren:Ang$_{mutated}$ model 46
3.3 Molecular dynamics simulations 48
3.4 QM/MM calculations 49
3.4.1 Determination of the potential energy surface 49
3.4.2 Increase of the QM region 51

CHAPTER 4 - RESULTS AND DISCUSSION 53
4.1 Introductory note 53
4.2 MD simulation 53
4.2.1 Root Mean Square Deviation 54
4.2.2 Root Mean Square Fluctuation 54
4.2.3 Active site interactions 55
4.3 The catalytic mechanism of human renin 58
4.3.1 The reactants structure 58
4.3.2 The first reaction step 59
4.3.3 The second reaction step 61
4.3.3 The third reaction step 62
4.3.4 Summary of human renin reaction 63
4.4 The role of a structural water molecule 64
4.4.1 The reactants structure 65
4.4.2 The first reaction step 65
4.4.3 The second reaction step 66
4.4.4 The third reaction step 68
4.5 Human renin catalytic mechanism with a mutated substrate 70
4.5.1 The reactants structure 71
4.5.2 The first reaction step 71
4.5.3 The second reaction step 72
4.5.4 The third reaction step 73
4.6 Energies associated with human renin mechanistic pathway 75
4.6.1 Energies along the reaction mechanism 75
4.6.1.1 Ren:Ang$_{dodecapetide}$:W1 model 75
4.6.1.2 Ren:Ang$_{dodecapetide}$:W1:W2 model 76

4.6.1.3 Ren:$Ang_{Mutated}$ model 77
4.6.2 Single Point energy calculations 78
4.6.2.1 Single point calculations using electrostatic embedding scheme 78
4.6.2.2 Single point calculation using large QM regions 79
4.6.2.3 Single point calculation using a large basis set 83
4.7 Human Renin vs. Mouse Renin 85

CHAPTER 5 - CONCLUSIONS AND FUTURE PERSPECTIVES 87
5.1 Conclusions 87
5.2 Future Perspectives 88

References 91

List of Figures

Figure 1. Classical renin-angiotensin system 4

Figure 2. Human renin structure complexed with its substrate....12

Figure 3. General proposal mechanism of amine bond hydrolysis by aspartic proteases. 14

Figure 4. The Lennard-Jones potential as a funtion of interatomic distance. 25

Figure 5. Representation of all interactions present in the potential energy expression....26

Figure 6. Representation of the ONIOM two-layer method....42

Figure 7. Representation of the different models used in the present work. 47

Figure 8. Representation of the system used in Molecular Dynamics Simulation. 48

Figure 9. Representation of the system used in QM/MM calculations 50

Figure 10. Representation of the atoms included in the different QM regions. 52

Figure 11. Renin and Renin:dodecapeptide RMSd along the MD simulation 54

Figure 12. RMSf of Renin and Angiotensinogen residues along the MD simulation. 55

Figure 13. Representation of the main active site interactions throughout MD simulation. 56

Figure 14. Representation of the renin:angiotensinogen system before and after the MD simulation. 57

Figure 15. Reactants, TS1 and INT1 obtained geometries for Ren:Ang$_{dodecapeptide}$:W1 model. 60

Figure 16. TS2 and INT2 geometries obtained for Ren:Ang$_{dodecapeptide}$:W1 model. 62

Figure 17. TS3 and Products geometries obtained for Ren:Ang$_{dodecapeptide}$:W1 model. 64

Figure 18. Reactants, TS1 and INT1 obtained geometries for Ren:Ang$_{dodecapeptide}$:W1:W2 model 67

Figure 19. TS2 and INT2 geometries obtained for Ren:Ang$_{dodecapeptide}$:W1:W2 model. 68

Figure 20. TS3 and Products geometries obtained for Ren:Ang$_{dodecapeptide}$:W1:W2 model. 69

Figure 21. Reactants, TS1 and INT1 obtained geometries for Ren:Ang$_{Mutated}$:model. 72

Figure 22. TS2 and INT2 geometries obtained for Ren:Ang$_{Mutated}$:W1 model....73

Figure 23. TS3 and Products geometries obtained for Ren:Ang$_{Mutated}$:W1 model. 74

Figure 24. Energetic pathway for angiotensinogen hydrolysis - Ren:Ang$_{dodecapeptide}$:W1 model 76

Figure 25. Energetic pathway for angiotensinogen hydrolysis – Ren:Ang$_{dodecapeptide}$:W1:W2 model....77

Figure 26. Energetic pathway for angiotensinogen hydrolysis - Ren:Ang$_{Mutated}$ model....78

Figure 27. Single point calculation with B3LYP 6-31G(d) level and electrostatic embedding scheme. 79

Figure 28. Single Point QM/MM calculations for a model with 55 atoms in high layer. 80

Figure 29. Single Point QM/MM calculations for a model with 100 atoms in high layer 80

Figure 30. Single Point QM/MM calculations for a model with 135 atoms in high layer 81

Figure 31. Single Point QM/MM calculations for a model with 210 atoms in high layer 81

Figure 32. Single Point QM/MM calculations for a model with 322 atoms in high layer 82

Figure 33. Single Point QM/MM calculations with a large basis set (6-311++G(2d,2p)) 84

List of equations

Equation 1. Force field potential energy 23

Equation 2. Bond Stretching potential energy 24

Equation 3. Bond bending potential energy 24

Equation 4. Dihedral angles potential energy.....24

Equation 5. van der Waals potential energy 25

Equation 6. Electrostatic potential energy.....26

Equation 7- Postulate of quantum mechanics.....29

Equation 8. Time-dependent Schrödinger equation 29

Equation 9. Time-independent Schrödinger equation 30

Equation 10. Hamiltonian operator expression 31

Equation 11. The Born-Oppenheimer approximation. 32

Equation 12. Slater determinant 33

Equation 13. The Hartree-Fock equations.....33

Equation 14. Fock operator expression 34

Equation 15. Energy functional equation (DFT) 35

Equation 16. Konh and Sham formalism 36

Equation 17. Exchange-correlation energy 36

Equation 18. Hybrid QM/MM energy 41

Equation 19. ONIOM total energy 42

List of tables

Table 1.Similarities and differences between Renin Inhibitors ACE inhibitors and AT1 receptors antagonists. 10

Table 2. Representation of the main interactions between the active residues and the residues around the active site during the reaction mechanism for Ren:$Ang_{dodecapeptide}$:W1 model. 64

Table 3. Representation of the main interactions between the active residues and the residues around the active site during the reaction mechanism for Ren:$Ang_{dodecapeptide}$:W1:W2 model. 70

Table 4. Representation of the main interactions between the active residues and the residues around the active site during the reaction mechanism for Ren:$Ang_{dodecapeptide}$:mutated model 74

Abbreviations

ACE – Angiotensin Converting Enzyme

AMBER – Assisted Model Building with Energy Refinement

Ang I – Angiotensin I

Ang II – Angiotensin II

AT receptor – Angiotensin receptor

DFT – Density Functional Theory

GGA – Generalized Gradient Approximation

GTO – Gaussian Type Orbital

H-GGA – Hybrid Generalized Gradient Approximation

HM-GGA – Hybrid Meta Generalized Gradient Approximation

HF – Hartree Fock

INT – Reaction Intermediates

K_{cat} – First order rate constant

K_M – Michaelis-Menten constant

LCAO – Linear Combination of Atomic Orbitals

LDA – Local Density Approximation

M-GGA – Meta Generalized Gradient Approximation

MD – Molecular Dynamics

MM – Molecular Mechanics

ONIOM – Our own N-layered Integrated molecular Orbital and molecular Mechanics

P - Product

PDB – Protein Data Bank

PES – Potential Energy Surface

PGTO – Primitive Gaussian Type Orbital

QM – Quantum Mechanics

QM/MM – Hybrid Quantum Mechanics and Molecular Mechanics

R - Reactants

RAS – Renin Angiotensin System

RMSD – Root Mean Square Deviation

RMSF – Root Mean Square Fluctuation

SCF – Self Consistent Field

SP – Single Point

STO – Slater Type Orbitals

TS – Transition State

Amino acid abbreviations

Ala – Alanine

Arg – Arginine

Asn – Asparagine

Asp – Aspartate

Cys – Cysteine

Gln – Glutamine

Glu – Glutamate

Gly – Glycine

His – Histidine

Ile – Isoleucine

Leu – Leucine

Lys – Lysine

Met – Methionine

Phe – Phenylalanine

Pro – Proline

Ser – Serine

Thr – Threonine

Trp – Tryptophan

Tyr – Tyrosine

Val – Valine

CHAPTER 1

INTRODUCTION

"Silent killer, public health crisis"

World Health Organization (WHO)

A global brief on hypertension – 2013

1.1 Introductory note

This work is being developed as part of a project with the title "New drugs for hypertension". One of the initial goals of this project is the study of the enzyme renin as a target for the control of hypertension. As such, this dissertation will describe the catalytic mechanism of this enzyme.

In this introductory chapter a brief review of the literature on some key topics useful for this work, will be made. Multiple subjects will be discussed, since the blood pressure, renin-angiotensin-aldosterone system, to the antihypertensive therapy currently used. A special attention will be given to the crucial biological role of the enzyme renin in blood pressure control, its structural analysis, and its mechanism as an aspartic protease. Renin will be analyzed according to a systemic and biological vision and subsequently by a molecular and atomic point of view. Also in this chapter the enzymatic catalysis and the rational development of new drugs will be addressed, as well as the importance of Theoretical and Computational Chemistry as an essential tool for their study.

1.2 Hypertension

Hypertension, or high blood pressure, is an important worldwide public health challenge because of its high incidence (more than 25% of adults worldwide) and association with risks of cardiovascular and kidney diseases. It has been identified on the ranking of higher risk factor for mortality and a disease that causes disability-adjusted life-years [1]. Hypertension is a chronic medical condition in which blood pressure in the vessels is elevated. The term blood pressure usually refers to the arterial pressure of the systemic circulation and it can be summarized by two measurements, systolic (when muscle's heart is contracted) and diastolic (when muscle's heart is relaxed), which corresponds to a maximum and a minimum pressure, respectively. Hypertension is said to be present when the blood pressure is persistently above 140 mmHg for maximum pressure and above 90 mmHg for minimum pressure. The statistic studies predict that in 2025 about 1.5 billion of people will have hypertension, which corresponds to an increase of 60% on the actual number. This increasing prevalence of hypertension worldwide has contributed to the pandemic presence of cardiovascular diseases that are responsible for 30% of all worldwide deaths [5, 6].

Interventions that have proven effective in hypertension include weight loss, reduced intake of dietary sodium, moderate alcohol consumption, modification of eating habits and increased physical activity. However, sometimes these prevention measures are not sufficient, and secondary prevention efforts require a detection, appropriate treatment and correct control of hypertension. Furthermore, in spite of the enormous advances in antihypertensive drug therapies, the number of people with uncontrolled hypertension has continued to rise and, thus there is a huge need to improve these therapies [7, 8].

The increasing number of patients with hypertension is also attributed, in part, to the growing problem associated to risk factors such as obesity and diabetes. Despite its prevalence, only 5% of patients with hypertension have an identifiable cause [9].

Hypertension is a product of dynamic interactions between multiple factors: genetic, physiological, environmental and psychological factors. Autonomic nervous system and kidney are the major drivers of hypertension, and recently some reviews reveal that the immune system play also an essential role in the hypertension pathophysiology [10, 11]. Independent of the organ system involving in blood pressure elevation, it is clear that modifications in vascular function and structure are paramount in a physiological point of view. At the molecular level, abnormal signal transductions are responsible for modifying

cellular function involving different signaling pathways, such as phospholipase C, inositol phosphate, diacylglycerol, mitogen-activated protein kinase, tyrosine kinases /phosphatases, Rho kinases, transcription factors and NAD(P)H oxidase-derived reactive oxygen species. These important signaling pathways are stimulated by pro-hypertensive peptides like angiotensin II (Ang II) through their interaction with membrane associated G protein-coupled receptors [9, 12] . The Ang II molecule is the final product of the renin-angiotensin system (RAS) that is the main regulator system of blood pressure. The next section will discuss this system, its constituents, regulation and importance.

1.3 Renin-Angiotensin System

The RAS is the master regulator of blood pressure and fluid homeostasis. This is an endocrine system and it is activated by various signals including the reduction of blood pressure, decrease in circulating blood volume and plasma-sodium concentration [13].

As representing in Figure 1, in a summary view, this system is a multi-enzymatic cascade that starts with angiotensinogen hydrolysis. This protein is cleaved in two steps, first by renin and then by angiotensin converting enzyme (ACE), resulting in two sequential peptides, the angiotensin I (Ang I) and angiotensin II (Ang II).

The appropriate activation of the RAS is essential for preventing circulatory collapse and maintaining fluid balance. A deregulation or a persistent RAS activation leads to a blood pressure increase.

1.3.1 Renin, Angiotensinogen and Angiotensin I

The first step of RAS

The synthesis and release of renin are the key and early steps of the RAS cascade. In spite of the very early discovery of renin, in 1898, by Tigerstedt and Bergman [14], we only have recently gained a deeper understanding of the mechanisms responsible for its synthesis and release [15]. Human renin is a 340 residues protease and it can be secreted by two cellular pathways: 1) a regulated way with secretion of mature renin or 2) a constitutive way with secretion of pro-renin. Mature renin is secreted into the blood stream only from juxtaglomerular cells of the kidney, in response to different cellular stimuli: a decrease of arterial blood pressure detected by baroreceptors, a decrease in sodium

chloride levels in the nephron's ultra-filtrate or sympathetic nervous activity, acting through β_1 adrenergic receptors. On the other hand, the precursor of renin, pro-renin can be synthetized, not only in juxtaglomerular cells, but also in other tissues like adrenal, collecting duct, eyes, Müller cells, ovary, theca cells, uterus, myometrium, placenta, chronically cells, testis and submandibular gland.

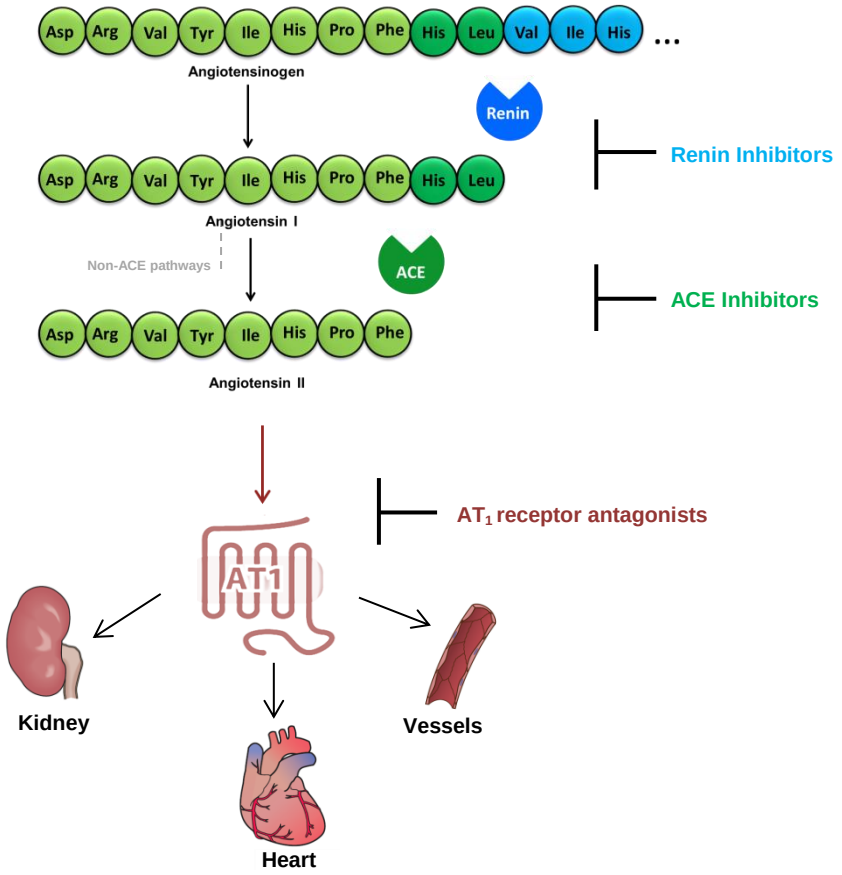


Figure 1. Classical renin-angiotensin system

Angiotensinogen, mainly expressed in the liver, is cleaved by renin to release Ang I in plasma. Then, ACE metabolizes Ang I to Ang II, which can also be converted by other non-ACE pathways. This peptide interacts mainly with the AT1 receptors in numerous organs such as kidneys, heart and vessels, and it is responsible for the increase of blood pressure. This system can be inhibited by renin inhibitors, ACE inhibitors or AT_1 receptor blockers (antagonists).

Renin is a member of the well-known family of aspartic proteases and it is responsible for the control of the first and rate-limiting step of RAS. This enzyme cleaves the 10 amino acids from the N-terminus of angiotensinogen, between Leu10 and Val11, to form Ang I.

Angiotensinogen is the only known physiological substrate for renin. In that way, renin has an essential and extremely specific role in RAS, in contrast to the ACE, as we will see later.

Human angiotensinogen is a 452 residues long peptide mainly produced in the liver. It is also produced by other tissues as brain, immune system and kidney, but in very small quantities. RAS is regulated by a negative feedback and, therefore, plasma angiotensinogen levels are increased by a decrease in the Ang II levels. In addition, plasma corticosteroid, estrogen, and thyroid hormones levels also promote an increase of angiotensinogen levels [16].

The cleavage of angiotensinogen by renin leads to the generation of both Ang I and des(Ang I)-angiotensinogen that corresponds to the remaining residues of angiotensinogen (more than 95%). This portion doesn't have any associated biological properties to date.

It is well known that renal renin is secreted into plasma to cleave angiotensinogen release from liver and no other function is associated with this enzyme. In addition to renin, other enzymes such as cathepsin D or G, pepsin, tonin, tripsin, have been demonstrated to convert angiotensinogen into Ang I. However, these enzymes are abundant only in specific tissues or in some specific cell types and only have this kind of activity under strict conditions. Therefore, they do not have any significant impact on the systemic production of angiotensinogen peptides [17].

The plasma concentration of angiotensinogen in humans is approximately 1 μM [18], which is close to the Michaelis-Menten constant (K_M) of renin (1.25 μM) [19]. This indicates that both angiotensin and renin concentrations are important for the rate of Ang I generation. The Ang I molecule, resulting from the first step of RAS, after the cleavage of angiotensinogen by renin, appears to have no biological activity. It exists only as a precursor to Ang II formation.

1.3.2 Angiotensin Converting Enzyme and Angiotensin II

The second step of RAS

ACE is a zinc dependent carboxyl dipeptidase. It is responsible for the cleavage of 2 residues from the C-terminus of inactive Ang I and it generates the Ang II, in the classical RAS axis (Figure 1).

The Ang II levels are also detectable in the ACE gene knockout mouse, which is indicative that enzymes other than ACE may contribute to Ang II formation. However, these studies also show that Ang II formation predominantly occurs via ACE [20, 21]. Along with Ang II, ACE also cleaves other peptide substrates, including the vasodilator peptide

bradykinin, which is degraded to an inactive peptide [22, 23] . Recently, ACE has also been implicated in the regulation of the immune system [24].

In the non-classical RAS system pathway another bioactive angiotensin, Ang-(1-7), and an ACE homologue, ACE2, add further complexity to the proteolytic cascade. This path is less efficient and it consists in a first hydrolysis of Ang I, by ACE, to form Ang-(1-9), a peptide that is not known to have biological activity. Then, this peptide can be converted to Ang (1-7). ACE2 can also cleave directly Ang II to form Ang-(1-7). This last peptide has important biological activities due to its cardiovascular and baroreflex actions counteract those of Ang II [22, 25, 26].

Although we cannot set aside the existence and importance of these new pathways, they will not be subjects of the present work. Ang II peptide is then the main and higher concentrated product of this second step of RAS. This is an active peptide that binds to the angiotensin receptors to perform their functions, as we will see in the next sub-section.

1.3.3 Angiotensin receptors

The third step of RAS

The actions of Ang II are mediated by specific populations of Ang II receptors. It is known that Ang II interacts at least with two distinct Ang II receptors subtypes, designed by AT_1 (Angiotensin receptors type 1) and AT_2 (Angiotensin receptors type 2). These subtypes of receptors can be distinguished and characterized according to inhibition by specific selective antagonists. There are also other subtypes of angiotensin receptors (AT_3 and AT_4) that have already some described functions. However, their activities and behavior are still poorly understood and characterized.

The main known biological actions of Ang II are the regulation of blood pressure, release of aldosterone, electrolyte and water balance, stimulation of sympathetic transmission, cellular growth. All of them are exclusively mediated by its interaction with the AT_1 receptor [27-29]. The AT_2 receptor, in turn, provides a counter regulatory role to AT_1 receptor over activity. Recent evidences suggest also that AT_2 receptor exhibits favorable organ protection in the context of the heart, kidney and brain, mediating anti-proliferation, apoptosis and differentiation [22, 29, 30].

When Ang II binds to the AT_1 receptor, which belongs to the G protein-coupled receptor superfamily, typically activates phospholipase C (coupled to a G_q protein) and leads to an increase in intracellular inositol 1,4,5-triphosphate and intracellular calcium. These intermediates are responsible for vasoconstriction effects and sodium and water retention. Downstream, protein kinase C, proteins as Ras and RhoA and tyrosine kinase cascades are amplified, affecting members of the mitogen-activated protein kinase family (ERK1/2 and p38) and JAK-STAT pathways leads to vascular and cardiac effects [31].

AT_1 receptors are located in the brain, adrenal cells, heart, vasculature and kidney. In blood vessels its activation by Ang II causes vasoconstriction leading to an increase in peripheral vascular tone and systemic blood pressure [29, 31, 32].

Contrarily to the AT_1 receptors, the signaling pathway mediated by AT_2 receptors leads to inactivation of ERK1/2, opening of potassium channels and inhibition of T-type calcium channel. Although various signaling pathways have been assigned to the AT_2 receptor, it still not clear which of these is the most important.

In summary, Ang II is a multifunctional hormone that exerts diverse physiological effects and their actions are mainly mediated by AT_1 and AT_2 receptors, leads to blood pressure effects.

1.4 Antihypertensive Agents

Drugs that inhibit RAS are gaining increasing popularity as initial medications for the management of hypertensive patients. Other antihypertensive agents are diuretics, beta adrenergic receptors blockers and calcium antagonists. Antihypertensive agents are used as monotherapy only in a small number of patients. The majority of people require two or more agents.

1.4.1 Diuretics

Diuretics, such as thiazides or related drugs, are some of the most frequently employed antihypertensive agents. The mechanism of action of these agents is related to the ability to increase sodium excretion as a result of binding to the thiazide receptor in the kidney. This therapeutic have, however, some side effects associated such as hypoatremia, hypokalemia, impotence and diabetes [33].

1.4.2 β-blockers

Other therapeutic agents are the β-blockers that have been used for more than 30 years. Some of the β-blockers are selective for β_1 adrenergic receptors, others are non-selective. Bradycardia, atrioventricular block or arrhythmias are normally associated to β-blockers. They may also adversely affect the lipid profile with increase of triglycerides in serum and are associated with diabetes. The mechanism of action of β-blockers may depend on the reduction in cardiac output [33-35].

1.4.3 Calcium channel blockers

The calcium channel blockers can also be used as antihypertensive agents. They act as peripheral vasodilators as they block the entry of calcium through voltage dependent calcium channels and they include dihydropyridines and non-dihydropyridines. These agents induce oedema, flushing and sometimes headache and palpitations [36] .

1.4.4 Vaccination

An old concept that has now resurrected is the use of vaccination to treat hypertension [37]. Some studies shows that a vaccine that induces antibodies against Ang II has yielded promising results in terms of blood pressure reduction [32]. However, the specificity and controllability of this kind of treatment needs to be evaluated.

1.4.5 RAS inhibition

The drugs that act in RAS system are also used as antihypertensive drugs. The pharmacological inhibition of classical RAS can be achieved through 3 different basic mechanisms, like we can see in Figure 1:

1) Inhibition of Ang I generation from angiotensinogen. This can be achieved by direct inhibition of renin.
2) Inhibition of Ang II generation from Ang I. This can be achieved through inhibition of ACE that, as we saw previously, is responsible for the formation of the octapeptide Ang II.

3) Inhibition of the action of Ang II at the level of its receptors.

1.4.5.1 Renin inhibitors

The most logical point to block pharmacologically the RAS is the rate limiting step catalyzed by renin enzyme. Intervention at this point is more specific than ACE inhibitor or AT1 antagonists, because almost any angiotensin peptide can be generated, and no other peptide system would be affected. The renin inhibitors, due to its advantages, have been studied and developed decades ago, but the first commercialized renin inhibitor, the aliskiren, came onto the market recently [38]. Aliskiren blocks the RAS at the origin of the cascade inhibiting the generation of Ang I, and consequently the generation of Ang II. As the main goal of this work is the study of the catalytic mechanism of renin, to allow future studies on its inhibition, it is essential to understand its role, and therefore more emphasis to this enzyme and its inhibitors will be given later.

1.4.5.2 ACE inhibitors

The first RAS inhibitor was an ACE inhibitor. Its intervention began in 1960's with the discovery of venom of the snake *Bothrops jararaca,* which contains a substance that inhibits ACE. By modelling the active site of ACE, several new drugs that potentially bind this site were designed, and the first orally available ACE inhibitor was captopril [39]. Today more ACE inhibitors are available on market. These drugs proved to have a strong beneficial effect on morbidity and mortality in congestive heart failure and renal diseases. As we mentioned before, ACE is also a bradykinin degrading enzyme, therefore its inhibition also increases bradykinin concentration that is a vasodilator peptide. It has also been show, by other studies, that bradykinin potentiation is involved in the cardio and renal protection. However, this peptide may also be the major reason for adverse side effects of these inhibitors, as cough and angioedema [33].

1.4.5.2 AT_1 receptor antagonists

A third group of drugs that interfere with RAS are specific antagonists for the AT_1 receptor. The first example of this class was losartan, which was followed by several other sartans. These drugs exert a more complete angiotensin blockade because other pathways

of angiotensin generation, such as cathepsins, tonin or chymase, that are not affected by ACE inhibitors, become ineffective by AT_1 antagonism. These drugs are also more specific for the RAS, than ACE inhibitors, because other peptide systems are not affected. The compensatory increase in renin concentration after AT_1 blockade leads to an accumulation of Ang II, which activates the AT_2 receptor. However, it is yet unknown whether this AT_2 stimulation often followed by bradkynin generation that is involved in the action of AT_1 antagonists [32, 40]. AT_1 blockers and ACE inhibitors work by different ways and both have effects on lowering the blood pressure, however, ACE inhibitors are currently best documented then AT_1 receptor antagonists. The choice of the best drug depends also on the patient.

The similarities and differences between renin inhibitors, ACE inhibitors and AT_1 antagonists are summarized on the Table 1. In this table, it is evident that the renin inhibition is the most rational target to control the RAS system and consequently the blood pressure. It is important to refer that by inhibiting ACE or blocking the AT_1 receptors, the negative feedback control, exerted by angiotensin II in the renin concentration, disappear and consequently an elevation on plasma and tissue renin activities are observed. Renin inhibition leads also to an increase on renin secretion and production but, in contrast to others RAS blockers, the newly release active renin is immediately neutralized by inhibitors present in the systemic circulation [41].

Table 1.Similarities and differences between Renin Inhibitors ACE inhibitors and AT1 receptors antagonists.
(+) symbol represents an increase of concentration and (-) symbol corresponds to a decrease of concentration. This table was adapted from [16].

	Renin Inhibitors	ACE Inhibitors	AT1 Antagonists
Plasma Renin Concentration	+	+	+
Plasma Renin Activity	-	+	+
Angiotensin I	-	+	+
Angiotensin II	-	-	+
Bradykinin	Unchanged	+	Unchanged
AT1 Receptor	Not stimulated	Not stimulated	Blocked
AT2 Receptor	Not stimulated	Not stimulated	Stimulated

Making a point of the situation of what has been described hitherto, the hypertension was initially introduced as a concern disease due to its direct relationship with cardiac and renal diseases. Then, the main regulatory system of blood pressure (RAS) was briefly

presented, with some emphasis to the first step. Finally, the most common antihypertensive therapeutics was summarized.

From now on, the renin enzyme will be the main focus of this work. Its structure, function and inhibition will be discussed, due to this enzyme appears to be the most logical target for the development of new antihypertensive agents.

1.5 Renin

1.5.1 Relation between structure and function

As previously mentioned the renin enzyme cleaves angiotensinogen to release the N-terminal decapeptide Ang I. Efforts into the elucidation of the molecular architecture of renin only succeed when they could be expressed and purified from recombinant sources.
Renin (EC 3.4.23.15) is a mono-specific 340-residues protease polypeptide, with approximately 40 000 KDa that belongs to aspartic protease superfamily [42].

Structurally, renin consists of 2 lobes with different residues sequence but similar structure. These lobes fold mainly β sheet conformation, with a deep cleft between them that is characteristic from the aspartic proteases class. The residues around the cleft correspond to the active site and it accommodates seven residues from its substrate, angiotensinogen *(***Figure 2**C). It has been shown that the minimal substrate sequence that has good affinity and reactivity is an octapeptide constituted by the residues between 6 and 13 of angiotensinogen. The catalytic activity of renin is due to two aspartic acid residues (32 and 215 or 38 and 236 in pepsin and human renin numbering, respectively), each located in a single lobe and held coplanar by a network of hydrogen bonds involving the surrounding main chain and conserved residues side chain groups. Like happens in other aspartic proteases, there are a water molecule in the active center of renin that hydrogen bonding with all four carboxyl oxygen atoms of aspartic catalytic dyad, and it has been implicated in catalysis. It has been suggested that this water molecule is partly displaced near the substrate and it is polarized by one of the catalytic aspartate residues.

The structure of renin has both hydrophobic and hydrophilic residues. The hydrophobic ones are more concentrated inside the protein and form hydrophobic cavities in the active site [41, 43, 44].

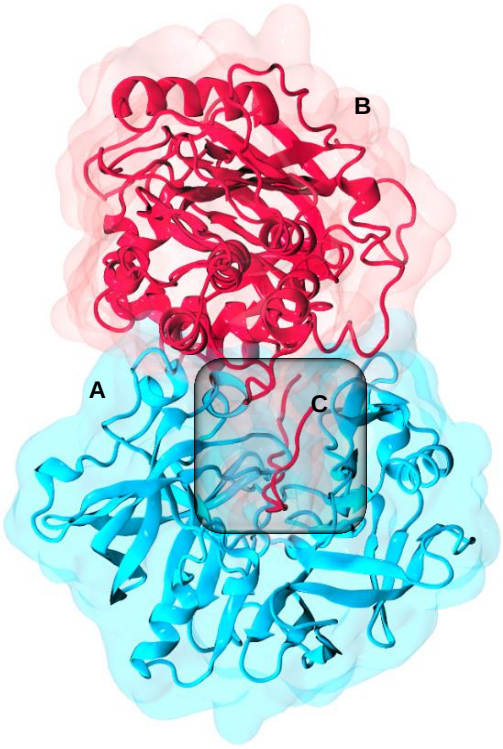


Figure 2. Human renin structure complexed with its substrate

A) Cartoon representation of human renin. B) Cartoon representation of human angiotensinogen interacting with the deep cleft of renin (C). The structure was adapted from 2X0B PDB file.

In the same way, as in the other aspartic proteases, renin possesses a flexible "flap" constituted by residues in a β-hairpin that is close to the active site and allows cleft covering. Recently, the dynamical behavior of this flap was studied by Molecular Dynamics (MD) simulations, and it was verified that it oscillates between open, semi-open and closed conformations could allow the binding of a wide range of inhibitors [45].

The active site of renin was studied by other authors and it is composed of 8 sub-pockets. The characteristics of each pocket was reviewed by Webb [46]. The S1 and S3 pockets forms a continuous and hydrophobic cavity and they contains the $S3^{sp}$ sub-pocket, which is specific to renin and unique among the aspartic proteases and have also hydrophobic and hydrophilic characteristics. The S2 and S4 pockets have also hydrophobic characteristics. The S1' are predominantly hydrophobic, while S2' have polar character (hydrophilic). This kind of information has a very important role for the future design of new renin inhibitors.

It is known that renin is very specific for its unique substrate and an important question that arises is that, if the main chains of proteases are conserved, how are the differences in

specificities achieved? Probably the specificities derive from slight differences in the sizes of residues in the specific pockets (S_n) to complement the corresponding side chain of the substrate. For example some studies have predicted the specificity site S_3 to be larger in renin than in other aspartic proteases and that a movement of a helix makes the pocket quite compact and complementary to aromatic rings. Therefore, the position of an element of secondary structure differs between renin and other aspartic proteases with important differences in the pocket specificity. As it happens in S_3 pocket, the differing positions of secondary structure elements may also account for the specificities of the other pockets. Previous studies were also suggested that this high specificity is due to the rigidity of its binding pocket. However, the reason of the high renin's specificity is still unknown [47, 48].

1.5.1 Proposed for Renin catalytic mechanism

The main objective of this work is to clarity with atomistic detail the catalytic mechanism of human renin by a computational approach. The mechanism of action of other aspartic proteases such as HIV-1 protease, presinilin and β-secretase was previously studied with MD simulations and combined Quantum Mechanics and Molecular Mechanics calculations (QM/MM) [49-53]. All of these computational techniques will be discussed in the following chapter.

There are proposals for the catalytic mechanism of renin but, its atomistic details remain unclear. Our group also published a proposal for the catalytic mechanism of mouse renin [4], but for the human enzyme, the mechanism has not been described yet.

The proposal mechanism of action of human renin involves an acid-base mechanism as it was described for other aspartic proteases. One of two catalytic aspartic residues is protonated, whilst the other, with a negative charge, acts as a general base, activating a water molecule tightly bound between both aspartate residues. Then, the water molecule makes a nucleophilic attack on the carbonyl carbon of the scissile bond resulting in the formation of a tetrahedral gem-diol intermediate (Figure 3A-B). The second step involves a protonation of the peptidic nitrogen by one of the catalytic aspartate and a deprotonation of one of the gem-diol hydroxyl group by the other aspartate. Subsequently, this step leads to the cleavage of the peptidic bond (Figure 3C-D).

In the beginning of the reaction the two catalytic aspartate residues have an adjacent position and they have different pK_a values, one being protonated and the other ionized at

neutral pH. This result was previously demonstrated by experimental and theoretical studies [54].

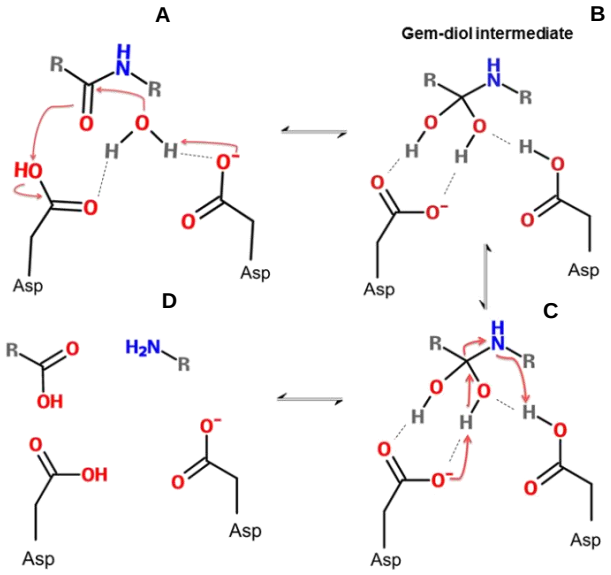


Figure 3. General proposal mechanism of amine bond hydrolysis by aspartic proteases.

A) Representation of the aspartic proteases active site with two Asp residues and a catalytic water molecule. In first step of the proposal catalytic mechanism a water molecule makes a nucleophilic attack on the carbonyl carbon of the scissile bond resulting in the formation of a gem-diol intermediate (B). C) In a second step, the nitrogen of the scissile bond is protonated by the catalytic Asp residue and this step ends with the peptidic bond cleavage (D).

It is also important to say that, in aspartic proteases, the residues around the catalytic center play an essential role in the catalytic mechanism. Furthermore a second conserved water molecule seems to have a structural role due to its participation in an H-bond network (Asp-Ser-W2-Tyr-Trp) that stabilizes the flap conformation during the catalysis [54]. However the role of this water molecule in the catalysis is still unknown.

1.5.3 Development of renin inhibitors

Like we mentioned before, the use of the drugs that inhibit the RAS system is an effective way to intervene in the control of hypertension. Considering the higher specificity of renin, the inhibition of this enzyme has benefits by decreasing its activity without interference with other metabolic pathways [55].

Pepstatin was the first renin inhibitor. It was found to be a competitive inhibitor of most aspartic proteases.

Then, considering that it was widely recognized that direct renin inhibition held great therapeutic potential, the design of new molecules that interact with the catalytic center of renin has been studied. First generation of renin inhibitors were peptide analogues (peptidomimetics) of the amino-terminal sequence of angiotensinogen. These analogues contain the renin cleavage site which, with further chemical modifications, led to the development of compounds as CGP29387 that had greater stability and long duration action, but it only have effect in high doses [56]. The second generation of renin inhibitors is characterized by a molecular weight of a tetrapeptide such as enalkiren, CGP38560A, ramikiren and zankiren. When given orally, these compounds had low bioavailability (< 2%), a short half-life and weak blood pressure lowering activity [57]. Later, the aliskiren was discovered as an orally active renin inhibitor by Ciba-Geigy (now Novartis). After successful preclinical and clinical testing it was approved and commercialized [58, 59]. Actually, with the help of crystallography and molecular modeling, the substituted piperidine renin inhibitors have been developed. However, the majority of them are abandoned after preclinical tests and no one was approved until now [55]. The Aliskiren, the unique drug commercialized for renin, interacts with several binding pockets in distinct regions around the active site. In particular it was found to bind the $S3^{sp}$ specific sub-pocket. This inhibitor has been shown, in experimental and clinical studies, to be effective in lowering the blood pressure and in organ protection. The Aliskiren effects are similar to those of other RAS inhibitors and normally it is used in combination with other antiherpentensive drugs.

Given the continuous prevalence of hypertension is still need to develop new drugs for this disease. With the advancement of computational techniques in enzymatic studies, the design of new effective drugs may be easier.

1.6 Computational enzymatic catalysis and drug design

The aim of this work involves the study of the catalytic mechanism of the renin enzyme. Therefore, a brief approach to enzymatic catalysis is also necessary in this introductory chapter. In fact, living cells and in turn multicellular organisms, depend on chemical transformations for every essential life processes. If these essential biochemistry reactions proceed at non catalyzed rates they are too slow (million years!) to sustain life. It is evident

that a very slow reaction cannot sustain life without some very significant rate enhancement. The enzymes are responsible for this enhancement and, therefore, enzymatic catalysis is essential for all life.

1.6.1 The role of computational chemistry in enzymatic catalysis

The first question to be asked is 'What is the catalytic mechanism of a reaction?' Establishing a mechanism of a reaction means identifying which groups in the protein are involved in the reaction, and clarifies their precise roles during catalysis. There are many experimental methodologies such as protein crystallography, NMR, site-directed mutagenesis, kinetics, isotopic labeling and spectroscopic those provide fundamental insights to the study of enzymatic catalysis. Research in the establishment of enzymatic mechanisms has suffered great advanced with new efforts in computational chemistry, which allows the atomistic description of all mechanistic details.

Predicting an enzymatic catalytic mechanism is generally a long and complex job. A complete description involves the identification of reactants (R), intermediates (INT), transition states (TS) and products (P) along the reaction coordinate, as well as the free energies of activation and reaction for each of the reaction steps.

In some cases, experimental methodologies allow the determination of the initial enzyme-substrate complex, the identification of some intermediates and the determination of the free energy of activation of the rate-limiting step. However, the structure of TS and the energies associated to them are impossible to determine by these techniques, because of the reduced lifetime of these species (10^{-13} seconds that corresponds to the time for a single bond vibration).

On the other hand, computational methodologies offer opportunity to study with atomic detail the enzymatic mechanism, providing atomistic insight in very important processes. All these facts indicate that computational studies have an essential and very important role in mechanistic studies. They are able to correctly identify and describe in a structural and energetic view the Transition states, independently of their lifetime, which complement the information obtained from experimental works [60-62]. In addition, computational methodologies make possible the study of large systems, like enzymes, where the reactive region is only a small part of the system. Thus the remaining structure of the enzyme and its role as a modulator of the active center, either stabilizing or destabilizing along the potential energy surface (PES) is also evaluated.

To study the correct mechanism of an enzymatic reaction, we need to search for the minimum energy pathway in the PES of the study system. The realism and success of the computational data for a mechanistic pathway depends on the system's Hamiltonian accuracy (the Hamiltonian will be explained latter), the sampling of the reactional space and the sampling of the conformational space [60, 63]. One of the key to elucidating the catalytic mechanism of an enzymatic reaction is the understanding of specific interactions between the substrate and its enzyme environment and their effect on the reaction rate.

1.6.2 Rational drug design

In drug discovery, enzymes are important as targets for new drugs (most part of drugs are enzyme inhibitors). The identification and description of enzymatic catalytic mechanisms add another dimension to structure based ligand design for enzyme targets. Modeling can also identify catalytically important interactions that could be exploiting in drug design.

All chemical enzymatic reactions pass through an unstable structure called the transition state (TS) with is poised between the chemical structures of the substrates and products. Enzymes stabilize the TS of their reactions. By binding the TS more tightly than the substrates or the products, the enzyme is responsible for lower the energy barrier of the reaction. Consequently a TS analog will bind more tightly to the enzyme than either the substrates or the products, preventing them from binding to the enzyme and reacting.

For aspartic proteases, like HIV-1 protease, there are many inhibitors that were obtained by drug design with TS analogs. Nowadays, HIV-1 protease inhibitors constitute around 40% of available drugs against HIV. Nearly all of them contain hydroxyethylene units as TS analog mimic. Thus, the design and success of HIV-1 protease inhibitors represent one of the most remarkable achievements of drug design [64].

Due to rapid technological progress in chemistry, bioinformatics, structural biology and computer technology, studies of computer aided drug design play more and more important role in this respect.

1.7 Purpose of the present work

As mentioned before, RAS represents the most important regulating system of the arterial blood pressure. The renin enzyme, an aspartic protease, has an exclusive activity

and its determinant velocity on the RAS cascade makes it an ideal target for new antihypertensive drugs. Therefore, the main purpose of the present work is describing with atomistic detail, the catalytic mechanism of this human renin by computational methods.

In addition, this study will also assess the role of a structural water molecule in the catalytic mechanism of this enzyme. Moreover, the role of residues around the active center, in the stabilization of the reaction intermediates, will also be evaluated. Finally, the influence of a mutated substrate, in the reaction catalyzed by renin, will also address.

This kind of knowledge has a very important role in future drug design which will enable the control of the “silent killer” and the “public health crisis” that is hypertension.

CHAPTER 2

THEORETICAL BACKGROUND

"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, (…) and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble."

Dirac, 1929

2.1 Introduction

Theoretical chemistry aims to develop algorithms, methodologies and theories that explain and reproduce systems at atomic and molecular levels. When they are applied computationally, the methods of theoretical chemistry, seek the clarification of chemistry and biological systems at an atomistic level.

If we wish to study the catalytic mechanism of an enzyme, as is the main purpose of this dissertation, we must use theoretical and computational methods. Computational biochemistry is able to predict results and deductions that are very expensive and time consuming when they are obtained experimentally. These methods are still able to interpret

experimental results and complement them, as well as suggest new interpretations for a better understanding of these results to clarify relevant doubts that may exist. Therefore, experimental techniques and theoretical techniques go hand in hand to try to explain chemical and biological problems at the molecular level.

Since the discovery of the resolution of the Schrödinger equation for the hydrogen atom in 1926 [65], quantum mechanics has been increasingly used in an attempt to interpret and predict the electronic structure and behavior of bio/chemical systems. A great number of efficient computer programs has been developed, which try to study with increasing accuracy large size systems, such as biological enzymatic systems. Due to the fast progress of computing power, the application of these methods to the study of biological systems has been widely used and the results obtained have been more and more precise. Through computational methods applied to Biochemistry it is possible to determine individual properties of biological molecules such as reaction energies, reaction mechanisms, dipole moments, charge distribution, interaction between different molecules among others. Thus, Computational Biochemistry is an extended scientific area that can be combined with any other area of Biochemistry, promoting the development of useful applications in Chemistry, Biology or Medicine. For example, the clarification of enzymatic reaction with computational methods can be extremely useful for the rational design of new molecules/drugs to receptors or enzymes in order to control, or even cure, some diseases.

Nowadays, Theoretical Chemistry methods are increasingly sophisticated and may be applied to more complex systems. Generally theoretical methodologies can be classified into four main categories:

1. *Ab initio* methods
2. Density Functional Theory (DFT)
3. Semi-empirical methods
4. Molecular Mechanics (MM)

Ab initio methods do not include any empirical parameters in their equations, and they are only based on the laws of quantum mechanics (QM). They try to solve the equation of Shrödinger as accurately as possible through a strict set of mathematical approaches. Those methods potentially provide very accurate results. However, they require a high computational cost, and therefore, they are only used for small size systems. **DFT** methods

are also quatum mechanical but they are bases on the total electronic density of the system. **Semi-empirical** methods maintain the quantum nature of the calculation, but, as the name suggests, parameters from experimental data can be introduced in the calculations. These methods are quite computationally inexpensive and rely on a series of approximations. **MM** methods are very useful when we have very large systems to study, as is the case of enzymes. In MM, the atom is the fundamental particle and classical mechanics laws are used to predict the behavior of the system, contrarily to the methods presented before. In that way, the MM force fields do not allow for an electronic description of the system and the description of bond breaking and forming is not possible. In addition to the methods presented previously, it is also possible to use **hybrid methods** in the study of large systems like enzymes. These include a combination of QM and MM methods, where the system is divided in two or more fractions and each part is calculated with one distinct method.

In the following sections the computational methods used in this work, as well as their development and application, will be reviewed briefly. The first section of this chapter will deal with the molecular mechanics and then another section will be dedicated to quantum mechanics

2.2 Molecular Mechanics

Molecular mechanics (MM) is a simplified approach to chemical problems. If we want to explore geometries, relative energies of conformers of the same molecule or study the behavior of docked substrates into active sites, MM methods are very effective tools. They are based on classical mechanics and they are very useful to perform calculations on systems containing a significant number of atoms [66-68].

In MM the smallest particle is the atom, which means that neither electrons nor protons are explicit considered and the potential energy depends only on the atom position. Molecules are treated as a set of charged point masses (the atoms) which are coupled together with springs. The total energy of a structure is calculated using an analytical function that sums the individual energy terms. Generally, the function includes bond stretching, bond angle bending, torsion, and non-bond interaction terms.

MM falls apart when one wants some deeper understanding of the system. If we are studying a reaction mechanism profile, breaking or forming bonds, we cannot use the MM

approach. The breakage and formation of bonds is out of scope of MM [69]. Despite some incapacities, MM is a very powerful method and it is extensively used in computational biochemistry as it allows the treatment of very large systems.

The parameters that are used in MM energy calculations are derived from databases of structures developed by experimental and quantum mechanical methods. These parameters and equations used to describe a system in MM are referred as a force field.

2.2.1 Basic Theory – The force field energy

The MM energy expression consists of a simple algebraic equation for the energy of a compound and it does not use a wavefunction or the total electron density. The constants used in these equations are obtained from experimental data or *ab initio* calculations, as referred previously. One of the ideas of molecular mechanics is the transferability of parameters. In other words, transferability hypothesis suggests that the properties of atoms in large molecules can be deduced from the study of representative set of small molecules. The derived energy parameters are then transferred to proteins or other larger and more complex molecular systems. Roughly speaking, the transferability implies that the bond lengths, angles (etc.) are the same in small "test" molecules as well as in much larger proteins. In order for the transferability of parameters to be a good description of the system, force field use atom types. This means that a sp^3 carbon will be described by different parameters than a sp^2 carbon, and so on [70].

Since MM does not take into consideration properties that depend on electronic distribution, several valid assumptions have to be taken into account. The Born-Oppenheimer approximation makes it possible to express the Hamiltonian as a function of the nuclear variables, since the motion of the electrons is averaged out. Therefore the classical potential energy tries to describe as accurately as possible the energy of the system using a rather simple model of the interactions within a system from processes such as stretching of bonds, the bond angle bending, dihedral torsion terms, the rotation around single bonds (intramolecular forces) and non-bonding interactions within a system with contributions such as Coulombic and Van der Waals interactions.

Therefore, the potential energy function can be described as a mathematical equation that allows for the potential energy of a chemical system to be calculated as a function of its three dimensional structure (Equation 1*)*. The potential energy of a system can be written as

a sum of terms (force field) that describe the energy required for the molecule to move and interact with its environment in a specific manner:

$$V(r^N) = V_{bonds} + V_{angles} + V_{dihedrals} + V_{van\ der\ Waals} + V_{Coulomb}$$

Equation 1. Force field potential energy

Where V_{bonds} represents the potential energy required for stretching a bond between two atoms, V_{angles} represents the potential energy function for angle bending, $V_{dihedrals}$ is the torsional energy for dihedral angles. Finally, the terms $V_{van\ der\ Waals}$ and $V_{Coulomb}$ describe the non-bonded atom-atom interactions. The $V_{van\ der\ Waals}$ describes the energy of the repulsion and attraction between atoms that are not directly bonded, and the $V_{Coulomb}$ represents the electrostatic potential energy between all atom pairs separately by three or more bonds.

A MM force field gives to each atom an atom type that depending on the atomic number and its molecular environment. Different force fields have different atom types. One of the most popular force fields is AMBER (Assisted Model Building with Energy Refinement) [71] within the AMBER software, which was used in the present work and it is essentially used for describing polypeptides, proteins and nucleic acids. There are other force fields specific for biomolecules such as CHARMM (Chemistry at HARvard Molecular Mechanics) [72] and OPLS-AA (Optimized Potential for Liquid Simulations) [73]. In addition, there are different force fields that possess parameters to describe other type of molecules such as carbohydrates and lipids.

In the next subsections, each term of the force field potential energy equation (Equation 1) will be summarized.

2.2.1.1 The Stretch bond energy

The potential energy of bond stretching corresponds to the energy associated with the elongation of a bond. It is given as a sum of potential energies for stretching each bond within the molecule under study. The simplest approach to calculate it is to use a harmonic potential centered at the equilibrium bond length (Equation 2).

$$V_{bonds,ij} = \sum \frac{1}{2} k_{IJ}(l_{ij} - l_{IJ}^0)^2$$

Equation 2. Bond Stretching potential energy

The expression takes V_{bonds} as a quadratic function of the displacement of the bond length l_{ij} from its equilibrium bond length l_{IJ}^0. In this expression, *I* and *J* represent the atom types of the atoms *i* and *j* in the molecule. The k_{IJ} constant and l_{IJ}^0 length depend on the atom types that form the bond and they are contained in the force field parameters.

2.2.1.2 The bending energy for angles

The bending energy is the energy required to modify an angle past its equilibrium point. The potential energy of angle bending is taken as a sum of potential energies for bending each angle in the system, which is represented in the expression of Equation 3.

$$V_{angles,ijk} = \sum \frac{1}{2} k_{IJK}(\theta_{ijk} - \theta_{IJK}^0)^2$$

Equation 3. Bond bending potential energy

Where the θ_{ijk} represents the angle between the atoms *i*, *j* and *k*, and θ_{IJK}^0 is the reference value for the bond angle type *IJK*.

2.2.1.3 The dihedral / torsion energy

The dihedral / torsion potential energy is taken as a sum of the terms $V_{dihedrals,ijkl}$ over all atoms separated by 3 bonds. The expression that represents this term is described below:

$$V_{dihedrals,ijkl} = \sum \frac{V_n^{ijkl}}{2}(1 + \cos(n\omega^{ijkl} - \gamma))$$

Equation 4. Dihedral angles potential energy

Where ω is the dihedral or torsional angle defined between the sequence of *IJKL* atoms. The dihedral angle is defined as the angle between bonds *IJ* and *KL* when they are

projected into the plane bisecting the *JK* bond. In the Equation 4, n represents the number of minima over 360º of the torsional potential and γ determines the location of the minima.

2.2.1.4 The van der Waals energy

The van der Waals potential energy, $V_{van\ der\ Waals}$, is a non-bonded term and it is used for atom pairs, *i* and *j*, separated by three or more bonds. This term describes the repulsion and attraction between atoms that are not directly bonded.

The $V_{van\ der\ Waals}$ term is usually a sum of interactions involving all possible 1 and 4, 1 and 5, 1 and 6 (etc.) atoms. It may be represented by the Lennard-Jones potential (Equation 5).

$$V_{van\ der\ Waals} = \sum_{1.\geq 4} V_{van\ der\ Waals,ij}$$

$$V_{van\ der\ Waals,ij} = 4\varepsilon_{IJ}\left[\left(\frac{\sigma_{IJ}}{R_{ij}}\right)^{12} - \left(\frac{\sigma_{IJ}}{R_{ij}}\right)^{6}\right]$$

Equation 5. van der Waals potential energy

The AMBER force field takes the $V_{van\ der\ Waals,ij}$ as a Lennard-Jones potential (Equation 5) [74], in which R_{ij} is the distance between the atoms *i* and *j*, ε_{IJ} is the minimum potential energy attainable and σ_{IJ} is the distance when the energy is zero. Figure 4 shows the graphical representation of the Lennard-Jones potential function for a better understanding.

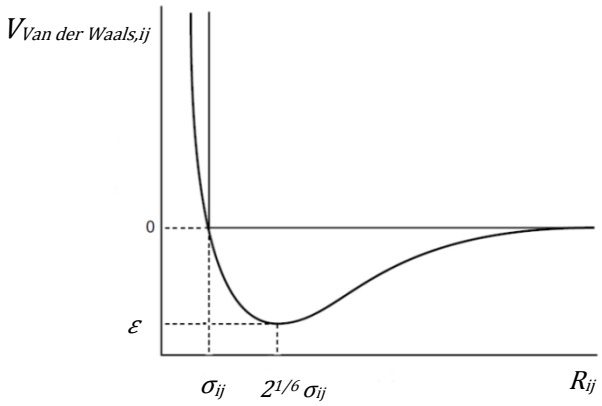


Figure 4. The Lennard-Jones potential as a funtion of interatomic distance.

The figure is adapted from *Essentials of Computational Chemistry-Theories and Models [67]*.

The Lennard-Jones potential contains a repulsive term, $\left(\frac{\sigma_{ij}}{R_{ij}}\right)^{12}$ which accounts for the electron repulsion. The 12th power dependence has no theoretical justification, and the 10 exponential factor produce, indeed, better results. However, the computational calculation is faster in that way because that term can be calculated by squaring the second term with exponential 6 [67]. The second term, $\left(\frac{\sigma_{ij}}{R_{ij}}\right)^{6}$, represents an attractive part that corresponds to the London's dispersion forces, in other words, the instantaneous dipole – induced dipole interactions. In summary, the Lennard-Jones potential is used to describe dispersion forces between separated atoms.

2.2.1.3 The electrostatic energy

The other term that also represents the non-bonded interactions is the electrostatic potential energy. The electrostatic potential can be written as a sum of electrostatic interactions involving all atom pairs separated by three or more atoms (Equation 6).

$$V_{Coulomb} = \sum_{1,\geq 4} V_{Coulomb,ij}$$
$$V_{Coulomb,ij} = \frac{q_i q_j}{4\pi\varepsilon_0 R_{ij}}$$

Equation 6. Electrostatic potential energy

The electrostatic potential between atoms *i* and *j* is described by a Coulombic term, where the q_i and q_2 represent the charges of the two atoms, the R_{ij} represents the distance between them and ε_r represents the dielectric constant of the medium.

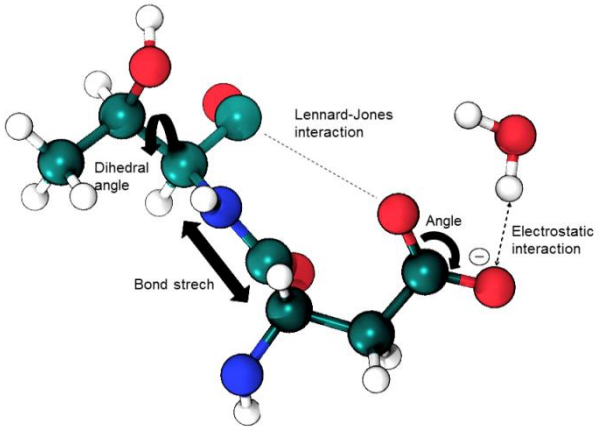


Figure 5. Representation of all interactions present in the potential energy expression (Equation 1).

2.2.2 Molecular dynamics

Molecular Dynamics (MD) simulates the motion of system particles as a consequence of their interaction with other particles.

Nowadays, the vision of proteins and nucleic acids as dynamic systems replaces the previous their view as static models. In that way, MD simulations are very important in the study of biological systems, because they are able to simulate the motion of some atoms that have essential biological functions.

The MD method was first introduced by Alder and Wainwright [75] in the late 1950s. The first protein MD simulation appeared in 1977, with the simulation of the bovine pancreatic trypsin inhibitor [76]. After this 1st application on this small protein, advances in computer power, algorithmic developments and improvements in the accuracy of the used interaction functions, have established this method as an important and predictive technique to study dynamic processes at atomic resolution. MD simulations can provide the ultimate detail concerning individual partial motions as a function of the time. They can be used to clarify specific questions about the molecular properties of proteins and nucleic acids (structure, dynamics and thermodynamics), often more easily and atomistic detailed than experimental methods [77].

In this method the simulation of a molecular system with the time, is given by the potential energy function presented previously, and the Newton second law or equation of motion. From the knowledge of the force on each atom of the system and its mass, it is possible to determine the acceleration of each atom in the system. Therefore, in molecular dynamics, the integration of the equations of motion, results in a trajectory that describes the positions, velocities and acceleration of each particle along time. When the position and velocities of each atom are known, the state of the system can be predicted at any time.

To calculate a molecular dynamics trajectory, it is just necessary to know the initial positions of the atoms, an initial distribution of velocities and the acceleration. The initial position of the atoms is given by experimental structures obtained, for example, by X-ray crystallography. The velocities are assigned randomly from a Maxwell-Boltzmann distribution at the desired temperature. The other parameters are determined by the potential energy function.

In order to obtain a satisfactory integration of the equations of motions, an adequate time step for the simulation must be chosen. To prevent some physical errors, the time step,

in which the forces are calculated, must be ten times smaller than the fastest process in the system. The fastest process in a simulation is normally the vibration of hydrogen atoms, which occur with a frequency of about 10 femtoseconds (fs) and a time step in the order of 1 fs are necessary to describe these motions.

In protein molecular dynamics studies, it is common to precede the simulation using explicit solvent molecules. The most typical water model used in the simulations is the TIP3P. The use of explicit water molecules brings some complications. While the atoms in the structure of protein are surrounded and bound by other atoms forming a cohesive structure, the particles near the boundaries (water molecules) can suffer large motions during the simulation. To solve this, a concept of periodic boundary conditions is employed, in which the main cell (protein, counter-ions and water molecules) is the central part of an infinite repetition of the identical cells in all directions. The movement calculated for the atoms in the original cell is replicated for the other cells. Thereby, the periodic boundary conditions are an efficient way to correct surface effects.

2.3 Quantum Mechanics

This subchapter deals with quantum mechanics and it will try to describe how its concepts can be useful in the enzymatic catalysis studies.

Until the end of the nineteenth century, physics consisted essentially of classical mechanics, electromagnetism and thermodynamics. Classical mechanics was used to predict the behavior of the dynamics of material bodies and the study of radiation was provided by Maxwell's electromagnetism. Matter and radiation were described in terms of particles and waves.

However, with Einstein relativistic theory and the development of the microscopic domain, the incontestable classical physics failed in providing the appropriate explanation for atomic and subatomic structures. Therefore, it became evident that the theories of classical physics cease at the microscopic level and new ideas had to be theorized in order to describe the structure of atoms and molecules. Many theories by Einstein, Planck, Bohr, Compton and De Broglie led to the definition of quantum mechanics. Quantum mechanics describes in an exact way the dynamics of all particles.

The wavefunction is the starting point for any discussion of quantum mechanics. The fundamental postulates of quantum mechanics describe microscopic systems as wavefunctions that completely characterize all physical properties of the systems. With the application of appropriate operators to the wavefunction, it is possible to determine observable properties of the system. The mathematical notation of this postulate is represented by Equation 7.

$$\vartheta\Psi = e\Psi$$

Equation 7- Postulate of quantum mechanics

Where ϑ is an appropriated operator for some property of the system, e is a eigenvalue this property and Ψ the wavefunction.

2.3.1 The Schrödinger equation

Mathematically, the Schrödinger equation is a function of an infinite space that maps all the possible states of a system. The time-dependent Schrödinger equation is represented by Equation 8:

$$\hat{\mathrm{H}}\Psi\,(q,t) = i\hbar\frac{\partial}{\partial t}\Psi(q,t)$$

Equation 8. Time-dependent Schrödinger equation

Where $\hat{\mathrm{H}}$ is the Hamiltonian operator, Ψ is the wavefunction, *i* is the imaginary unit, $\hbar$ is h / 2π, *q* are the particle coordinates and *t* is time. With this simple expression it is possible to calculate the evolution of any system along the time just by applying the $\hat{\mathrm{H}}$ operator to the Ψ.

With the solutions of the Schrödinger equation it is possible to describe molecular, atomic and subatomic systems. In addition it is also possible to describe macroscopic systems or, in an ideal vision, the whole universe. Although this is an exciting idea, it does not solve all chemical problems. The present equation is not analytically solvable for more than two interacting particles. As Dirac said, in the transcript that we use at the beginning of this chapter:

> "…*the exact application of these laws leads to equations much too complicated to be soluble."*
>
> *Dirac, 1929*

Many approximations are required to make this equation solvable. In the non-relativistic case, the energy of a stationary state of a given number of atoms and electrons is given by the time independent Schrödinger equation.

$$\hat{\mathrm{H}}\Psi = E\Psi$$

Equation 9. Time-independent Schrödinger equation

In this expression the solutions and the total energy of an isolated system are constants and no parameters are time dependent. The Hamiltonian operator returns the system energy.

This form of the Schrödinger equation also neglects relativistic effects. According to the theory of relativity, the mass of any particle has a significant change when this particle moves at velocities near the speed of light. This effect is only significant for heavier atoms (actinides and lanthanides) hence the inclusion of a relativistic term in the Hamiltonian is not necessary for biological studies. The time independent Schrödinger equation (Equation 9) is only a function of space coordinates (q) of the particles of the system.

2.3.2 The Hamiltonian

In quantum mechanics, the operators act on wavefunctions resulting into the same wavefunctions and the value of the physical properties that are associated with them. The Hamiltonian is an operator and it is associated with the total energy of the system.

In the time independent Schrödinger equation (Equation 9) we treat stationary states (constant energy) and as a result, Ψ is a function that, for atoms/molecules, only depends on nuclear and electronic positions.

The Hamiltonian operator is given as a sum of kinetic ($\hat{T}$) and potential energy ($\hat{V}$) operators and, for atoms/molecules, it can be written as follows:

$$\hat{H} = \hat{T} + \hat{V} = -\sum_{A}^{N} \frac{\hbar^2}{2m_A} \nabla_A^2 - \sum_{i}^{M} \frac{\hbar^2}{2m_e} \nabla_i^2 + \sum_{A<B}^{N} \frac{Z_A Z_B}{r_{AB}} - \sum_{A}^{N} \sum_{i}^{M} \frac{Z_A}{r_{iA}} + \sum_{i<j}^{M} \frac{1}{r_{ij}}$$

Equation 10. Hamiltonian operator expression

The Hamiltonian is written in atomic units and A and B refer to the *N* nucleus, *i* and *j* to the *M* electrons, and m_A and Z_A are the mass and atomic number of nuclei A, respectively. r_{AB} represents the distance between nuclei A and B, r_{iA} the distance between nuclei A and electron *i* and, finally the r_{ij} represents the distance between electrons *i* and *j.* The ∇^2 term is the Laplacian operator and is equal to $\nabla^2 = \frac{\partial^2}{\partial_x^2} + \frac{\partial^2}{\partial_y^2} + \frac{\partial^2}{\partial_z^2}$.

As we can see in Equation 10, the Hamiltonian has five terms: 1) the kinetic energy of the nucleus; 2) the kinetic energy of the electrons; 3) the coulombic energy for repulsions between the nuclei; 4) the coulombic energy for the attraction between nuclei and electron and 5) the coulombic energy for the repulsion between electrons. The potential energy terms (the last three) appear exactly as they do in classical mechanics. The kinetic energy is expressed by the kinetic energy operator: $\hat{T} = -\frac{\hbar^2}{2m} \nabla^2$.

Unfortunately, this expression cannot be solved exactly for more than two interacting particles at a time. Only single electron systems, where a single electron interacts with the nuclei, like the H atom or He^{1+}, Li^{2+} ions, can be treated with analytically. Systems with N ≥ 3 cannot be calculated exactly, because of the interaction terms that do not allow for a separation of variables [66]. In that way, it is necessary to introduce some simplifications and treat the systems in an approximate manner.

2.3.3 The Born-Oppenheimer Approximation

Up to now, we have been discussing many particle molecular systems entirely in the abstract. In fact, the wavefunctions for this kind of systems are extremely difficult to express because of the correlated motion of particles. In order to simplify the problem, the Born-Oppenheimer approximation [78] may be used.

Under physical conditions, the nuclei of molecular systems are heavier than the electrons and consequently they move much slower than the latter. In practice the electronic adaptation to a nuclear arrangement is very fast. Based on this hypothesis, it is possible to consider that the motion of the nucleus is not affected by the motion of electrons. Therefore,

the nuclear kinetic energy term can be calculated separately and the nuclear repulsion term can be considered as a constant for a given nuclear configuration.

In basic terms this approximation allows the wavefunction to be broken into two distinct equations: 1) an electronic Hamiltonian,$\widehat{H}_e$, with fixed nuclear positions and 2) a nuclear Hamiltonian, $\widehat{H}_n$, that is easy to solve (Equation 11).

$$\widehat{H}_{total} = \widehat{H}_{el} + \widehat{H}_n$$

$$\widehat{H}_{el} = \hat{T}_{el} + \hat{V}_{n-el} + \hat{V}_{el-el} \quad \text{(A)} \qquad\qquad \widehat{H}_n = \hat{V}_{n-n} + \hat{T}_n \quad \text{(B)}$$

Equation 11. The Born-Oppenheimer approximation.
A) Electronic contribution. B) Nuclear contribution

The calculation of the electronic energies for fixed nuclei means that the nuclear kinetic energy term is taken to be independent of the electrons, the correlation between electrons and the nuclei term is eliminated and the repulsive term between nuclei becomes constant for a given geometry. With this approximation the electronic Hamiltonian ($\widehat{H}_{el}$) includes only the second, third and fourth terms of Equation 10.

By using the Born-Oppenheimer approximation it is possible to extend the range of systems for which the Schrödinger equation could be solved exactly, *e.g.* systems with two nuclei and one electron, such as the H_2^+ ion. With this gain, this approximation greatly simplifies the calculation for any system.

2.3.4 Hartree-Fock Theory

Until now, we are able to solve analytically the energy of a system that contains two nuclei and a single electron, based on the Schrödinger equation and some approximations.

With the Hartree Fock Theory, any system can be treated. The upper limit in the number of particles that can be used is only a computational matter.

In the Hartree-Fock theory the correlation between electrons is discarded. This correlation is only included as an average effect. It is considering that each electron interacts with the effective field due to the nuclei and the average distribution of the other electrons. As mentioned before, the correlation factors between particles is the big problem that excludes the possibility to calculate analytically the energy of any system (with more than two particles). Hartree Fock calculation starts with generating a guess wavefunction

with the form of a Slater determinant (represents by Equation 12). On this expression the numbers represents electrons and the letters represents orbitals.

$$\Psi_{electronic} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_a(1) & \phi_a(2) & \dots & \phi_a(N) \\ \phi_b(1) & \phi_b(2) & \dots & \phi_b(N) \\ \dots & \dots & \dots & \dots \\ \phi_N(1) & \phi_N(2) & \dots & \phi_N(N) \end{vmatrix}$$

Equation 12. Slater determinant

The factor $\frac{1}{\sqrt{N!}}$, is used in order to normalize the wavefunction.

The employment of the variational principle is very important in the Hartree-Fock calculations. This determinant assurance that any approximate wavefunction corresponds to a higher energy than the real wavefunction. This insures that in order to improve the guess wavefunction, the energy of the system must be minimized.

A Slater determinant is composed of spin orbitals, *i.e.* the orbital function multiplied by its electron spin. This Slater determinant and spin orbitals are responsable to assuring that the Pauli principle is obeyed.

This principle states that the total wavefunction must be anti-symmetric under the interchange of any pair of fermions (as electrons). The interchange of two electrons is accomplished in a Slater determinant (by exchanging two columns), and was the consequence of inverting the sign of the wavefunction, according to the Pauli principle.

The principle that no two electrons can occupy the same quantum level (the Pauli exclusion rule) is also obeyed by this method. To guarantee that the Pauli exclusion principle is not violated, when multiple electrons occupy the same spin orbital, the rows corresponding to that spin orbital are equal and this leads to a null wavefunction.

Within the Hartree-Fock theory, the electronic Hamiltonian results in an operator for each electron, all electrons interact with a uniform electronic cloud, and the electrons are distributed by spin orbitals. Therefore, in order to obtain the energy, for each electron we have:

$$\hat{F}\phi_i = \varepsilon_i^{HF}\phi_i$$

Equation 13. The Hartree-Fock equations

where ϕ_i represents the wavefunction for the spin orbital of the electron *i* and ε_i^{HF} represent the energy associated with this spin orbital. $\hat{F}$ is the Fock operator that is described by the follow expression:

$$\hat{F} = \hat{h}_i + \sum_u \left\{2\hat{J}_u(1) - \hat{K}_u(1)\right\}$$

Equation 14. Fock operator expression

The Fock operator is a kind of Hamiltonian constructed individually for each electron. The $\hat{h}_i$ is the core Hamiltonian for one electron, which is composed by two terms: 1) the nuclei and electrons interactions and the kinetic energy of the electrons. The sum is over all other spin orbitals and it contains the operator for coulombic interactions between the electrons of the system $\hat{J}_u$, and the exchange integral operator, $\hat{K}_u$. These operators depend on other spinorbitals and, as a result, every Hartree-Fock equation is a function of one other. To solve this kind of equation a Self Consistent Field (SCF) calculation is done, with a guess Slater determinant and iterative optimizations, until and adequate convergence is achieved.

Although it has been widely accepted, the Hartree-Fock theory has only analytical resolution for atoms. Thus, in order to enable the study of molecules it was formulated a new strategy that will be briefly described in the following section.

2.3.5 Linear combination of atomic orbitals

In 1951, Roothaan and Hall proposed, for molecular orbitals, a wavefunction as a linear combination of atomic wavefunctions. The set of these wavefunctions is called the “basis set”. This construction is known as linear combination of atomic orbitals (LCAO) approach.

The combination gives rise to the molecular orbitals and consists of a predefined set of functions that describes atomic orbitals. The accuracy of the results depends on the number of basis functions used to describe the orbitals. It is important to say that the greater number of basis functions, the greater the computer time required. Therefore, it is necessary to get a balance and a compromise between the number of basis functions and the computating time.

2.4 Density Functional Theory

As seen before, the Schrödinger equation is exact, but impossible to solve for most relevant systems. The Density Functional Theory (DFT) is another set of methods used to elucidate the chemical and physical properties of a system that has appeared during the last decades as a powerful methodology. It introduces a different approach based on the calculation of the total electronic energy and electronic density distribution, without needing to guess the full N-electrons wavefunction as done in the previous method.

This method was introduced by Thomas and Fermi in 1927 [79, 80] and it is based on the principle that the energy of an electronic system can be defined in terms of its electron density, *ρ.* When a system has *n* electrons, *ρ(r),* represents the total electron density at a point of space *r*. The DFT formalism shows that the electronic energy *E*, is regarded as a functional of the electron density *E*[*ρ*]. It means that to a given function *ρ(r),* corresponds a single energy. The main advantage of a DFT method is that it depends only on three coordinates, independently of the number of electrons that constitutes the system, contrarily to the wavefunction calculated in the previous methods that depend on the three coordinates of each electron and one more if the spin is included (4n coordinates, where n is the number of particles) [68, 81].

2.4.1 The Hohenberg-Kohn Theorem

The DFT methodology is based on the Hohenberg-Kohn theorem [82]. This theorem was published in the 1960s and it affirms that in a non-degenerated system, the ground-state energy is uniquely defined by its electron density. In DFT the energy functional is written as a sum of terms as it is portrayed in equation 15:

$$E[\rho(r^N)] = \int V_{ext}\,(r^N)\rho(r^N)dx + F[\rho(r^N)]$$

Equation 15. Energy functional equation (DFT)

The first term is related to the electrons with an external potential ($V_{ext}\,(r^N)$) and usually representing the Couloumb interactions with the nuclei. The second term is the sum of the kinetic energy of the electrons and the electron correlation interaction, the correlation

energy. The main problem with this theorem is that the function $F[\rho(r^N)]$ is not described, it is unknown.

2.5.2 The Konh-Sham Theorem

Fortunately, in 1965, Konh and Sham [83], suggested a formalism that is in the foundation of the current application of the DFT methodology. Following this formalism the $E[\rho(r^N)]$ can be approximated as a sum of three terms and simplified as:

$$F[\rho(r^N)] = E_{KE}[\rho(r^N)] + E_H[\rho(r^N)] + E_{XC}[\rho(r^N)]$$

Equation 16. Konh and Sham formalism

where $E_{KE}[\rho(r^N)]$ is the kinetic energy of non-interacting electrons, $E_H[\rho(r^N)]$ is the electron-electron Coulombic energy, and $E_{XC}[\rho(r^N)]$ accounts for the contribution from exchange and correlation interactions.

As in the Hartree-Fock theory, to solve Hohenberg-Konh and Konh-Sham equations it is necessary to use a self-consistent approach. This process allows the Kohn and Sham equations to be solved, yielding an initial set of Kohn-Sham orbitals. This set of orbitals is then used to calculate an improvement of the density and the process is repeated until exchange-correlation energy have satisfied a certain convergence criteria, or, in other words, iterative steps, involving a variational process, necessary in order to describe the E_{XC} term.

In DFT, it is a common procedure to divide the exchange-correlation energy in two separate terms:

$$E_{XC}(\rho) = E_X(\rho) + E_C(\rho)$$

Equation 17. Exchange-correlation energy

where the $E_X(\rho)$ is the exchange functional associated with the interactions between electrons of the same spin and $E_C(\rho)$ is the functional of correlation between all electrons of the system.

The only approximation that needs to be done in the DFT methodology is due to the exchange correlation functional. The Local Density Approximation (LDA) functionals constitute the simplest approach to represent this exchange. It assumes that the exchange

correlation energy, at a point in space, is a function of the electron density at that point only, and can be given by the density of electrons in a homogeneous free electron gas. In that way, an electron feels the electron density produced by the remaining electrons as if the density was the same in each part of the system. However, for systems that present changes in electronic density in space, as is the case of biological molecules, the results from this approximation are unsatisfactory.

In order to improve the previous method, it is necessary to consider a non-uniform gas. In this case, the exchange and correlation functionals depend not only on the electronic density, but also on the derivatives of the density, and take into account the spatial varying density $\rho(r)$. This approximation is known as the Generalized Gradient corrected Approximation (GGA). A new class of functional based on GGA was developed and they are termed meta-GGA (M-GGA) methods. They depend explicitly on higher order density gradients or on the kinetic energy density as well, which involves derivatives of the occupied Kohn-Sham orbitals. Another different approach is the hybrid density functional (H-GGA). This method combines the exchange correlation of the GGA method with inclusion of a percentage of Hartree-Fock exchange. Some of the H-GGA also include constants that are calculated by mathematical adjustments through experimental results. An example of an H-GAA functional is the B3LYP, probably the most popular density functional used in computational chemistry, and also in the present work. Hybrid meta-GGA functionals (HM-GGA) represent a new class of density functionals. They are similar to the M-GGA functionals but they depend on the Hartree-Fock exchange, the electron density and its gradient, as well as on the kinetic energy density. These methods show an improvement over the previous formalisms, predominantly in the calculation of barriers heights and atomization energies. To conclude, the sequence of accuracy for density functionals is the following (from less to the most accurate): LDA<GGA<M-GGA<HM-GGA<Exact functional. However, it is important to say that from system to system, or from property to property, the performance of each functional can vary.

It is also important to refer again that the main advantage of the DFT methodology is the low computational cost and the inclusion of correlation effects, when compared with the Hartree-Fock methodology. The DFT methods can be applied to systems with few hundreds atoms and they are advantageous in relation to the methods presented previously.

2.5 Basis set

A basis set is a set of mathematical functions that are used to represent atomic orbitals in a quantum mechanical calculation. The larger the basis set, the fewer constraints are attributed to electrons and more accurate are the results. A complete basis set permits an optimal description of the electron probability density. However, in practice, a description of a complete basis set is impossible because it would involve an infinite number of functions. However, by finding an adequate set of basis functions, we can get good results. The mathematical functions used to mimic the atomic orbitals can be divided into several types and the most commonly used are the Slater type orbitals and the Gaussian type orbitals, which will be briefly described in this section.

2.5.1 Slater type orbitals

The Slater type orbitals (STO) describe the electron density around an atom. They have an exponential dependence on the distance of the electron to the nucleus that reflects the behavior of orbitals in the hydrogen atom.

The exponential dependence assures a simple convergence with the increase in the number of functions. However, the calculation of three or four centers integrals is not possible analytically. Therefore, the STO are particularly useful in mono or diatomic systems.

2.5.2 Gaussian type orbitals

In turn, in Gaussian-type orbitals (GTO), the most used type of basis set, the bi-electronic integrals are possible to calculate analytically even when the basis functions are centered in more than 2 atoms.

Unfortunately, with regards to the electron density, they are not able to describe it as accurately as STOs do. This happens specially when the distance from an electron to the core increases or decreases very much. Although exponential functions are a better fit to describe the electronic density, the GTOs are preferable compared to the STOs, because the first use simplified functions that reduce the computational cost. The GTOs can be improved by the use of a linear combination of Gaussian functions (called a contracted basis

set-CGTO). These Gaussian functions are called primitives (PGTOS) and have fixed coefficients. These functions improve the description of the atomic orbitals near the core. Moreover, as the electrons closer to the core have small chemical importance, this is a good approximation and it does not compromise the final results.

To improve the basis set described before, named “minimal basis sets” as each electron is described by a single CGTO, the split valence basis sets are used. They have a different scheme for core electrons and for valence electrons. The first ones are treated with a single contracted basis with *n* Gaussian primitives and the second ones are treated with two or three basis. The split valence basis were created by Pople reference and their nomenclature is represented by:

$$\boldsymbol{n-abc}\mathrm{G}$$

where n represents the number of PGTO used for the core orbitals and abc, represent the basis used to treat the valence part and their division. The G indicates that the basis used for describing the system were the Gaussian type orbitals. In this work, as we will see later, the 6-31G basis set was used. It is a double zeta basis set (two basis are used to describe the valence electrons) and it is one of the more used basis sets. In this case, 6 GTO primitives are included in the contracted core orbitals and 2 basis are used for the valence electrons. The more internals orbitals are described by 3 primitive Gaussian functions and the external ones are described by 1 primitive Gaussian function.

Despite the advantages, these basis sets have the disadvantage of being inflexible or, in other words, they do not change their shape with the molecular environment. This happens because this type of basis set does not describe orbital deformation when they are in a field created by nearby atoms. However, other methods have been applied in an attempt to provide greater flexibility to the wavefunction. The problem is that the s and p functions centered on the atoms do not provide sufficient mathematical flexibility to describe, in an adequate way, the wavefunction for different geometries.

To include flexibility in the Pople basis set there are two types of functions that can be added, the **polarization functions** and the **diffuse functions**. The first are almost always added in the form of basis functions corresponding to a higher angular momentum quantum number than the valence orbitals. This kind of basis sets are denoted by, (for example, for the case of 6-31G) 6-31G** or 6-31G(d,p). The first asterisk, or d, concerns heavy atoms which have d functions that polarize p functions. The second asterisk, or p, includes hydrogen atoms with s valence electrons that are polarized by p functions. It is important to

say that the 6-31G(d) is the most widely basis set used because it guarantes a good compromise between computational cost and accuracy. In this work this basis set was used in all optimization calculations.

The other kind of functions, the **diffuse functions**, are important to improve the description of electronic densities that are spatially diffused and far from the nuclei. When a basis set does not have a good flexibility to allow a weakly bound electron to localize far from the remaining density, significant errors in energies and other molecular properties can occur. Diffuse functions are useful in the case of anions, highly excited electronic states and large atoms that tend to have more diffuse electronic density than the others. The presence of diffuse functions is indicated by a "+" in the basis set name. If there are two plus signals in the basis set, the first one corresponds to the heavy atoms and the second one to the hydrogen atoms.

The choice of a correct basis set is an essential step in a calculation.

2.6 Hybrid QM/MM methods

We have seen in the previous section that it is possible to study chemical reactions from different perspectives.

For reasons of efficiency, the representation of large systems is most typically carried out at the MM level. However the molecular MM force fields, described before, are not sufficiently flexible to describe processes in which chemical bonds are broken or formed. To adequately describe such processes, QM methods are required [70]. Nevertheless, there are cases in which the systems are very large to treat with QM methods. To study enzymatic catalytic reactions it is important to include the presence of the enzyme surrounding. The smallest enzymes have thousands of atoms and therefore it is impossible to study this kind of systems only with QM methods.

It is known that the region of the enzymatic space in which significant changes in electronic structure occur, the active site, is relatively small compared to the size of the whole molecular system and, thus it can be treated with QM methods. The remainder of the enzyme may be important to maintain its structure and folding that are crucial for its activity, but it has small influence on the chemical reaction that takes place in the catalytic active site. In that way, from a modelling perspective, we may regard the situation within a limited

region, in which we should make use of quantum mechanics to accurately describe the electronic structure. In the surrounding region, the explicit representation of the whole system is important, but the level of model applied can be reduced in complexity. These methods, with a mixture of two or more theoretical levels are called *hybrid methods*.

In QM/MM the level applied to the outer system is MM and in that way, the complete energy for the system must be some kind of hybrid of QM and MM methodologies:

$$E_{complete} = E_{QM} + E_{MM} + E_{QM/MM}$$

Equation 18. Hybrid QM/MM energy

where E_{QM} accounts for full internal interaction energy of all quantum mechanical particles with one other, E_{MM} accounts for full internal interaction energy of all classical particles with one other and, finally, the $E_{QM/MM}$ accounts for all interactions between a quantum mechanical particle and a classical particle. Hybrid methods allow the combination of two or more computational techniques in one calculation and therefore it makes it possible to investigate large systems.

An important aspect to take into account when using hybrid methods is the division of the system. The partitioning of the system must involve the separation of covalently bonded atoms into different levels of theory. The way the system is "cut" must be made with special attention, because an inadequate choice of the atoms included in the QM or MM regions can lead to incorrect descriptions of the study system.

2.6.1 The ONIOM method

There are different QM/MM methods in which the division of the study system and the procedure for the energy calculation is different between them. In this work the QM/MM methodology used is called ONIOM method.

ONIOM (our own N-layer integrated molecular orbital molecular mechanics) [84-87] scheme can combine any number of molecular orbital methods, as well as molecular mechanics methods. In a two layer ONIOM calculation, the total energy of the system is obtained from three independent calculations:

$$E^{ONIOM} = E^{real\ system,MM} + E^{model,QM} - E^{model\ system,MM}$$

Equation 19. ONIOM total energy

The real system contains all the atoms and is calculated only at the MM level. The model system contains the part of the system that is treated at the QM level (high layer). Both QM and MM calculations need to be carried out for the *model system*. The schematic representation of the ONIOM model is shown in Fig.2.3. When there are bonded interactions between the two regions (QM and MM), the model system includes link atoms (in this case, hydrogen atoms) to saturate the open valence of the QM atom, as explained later [84].

A technical issue needs to be taken into account when we are studying chemical reactivity. When bond breaking and forming is part of the process in the QM region, it is safest to have the MM region at least three bonds away from the chemical process to avoid the cancelation problem. However, depending on the exact parameters involved in the MM terms in the border region there are instances where one or two bond separations are sufficient.

In the QM/MM methods there are two choices for dealing with the electrostatic interactions between the QM layer and the MM layer. The first, mechanical embedding treats the cross region electrostatic interactions at the MM level. The second, electrostatic embedding incorporates the cross region electrostatic interactions in the QM Hamiltonian.

The next subsection will give a brief explanation of these representations.

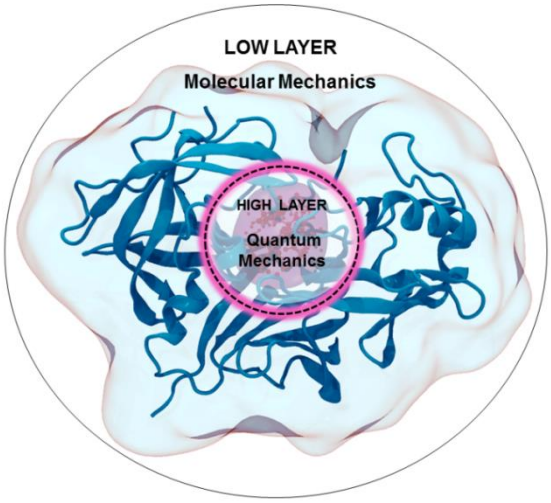


Figure 6. Representation of the ONIOM two-layer method

The system is divided in two layers. The High layer is treated with an accurate method, QM, and is the region where the catalytic reaction occurs. The remain system is treated with MM.

2.6.1.1 Mechanical embedding scheme

In the mechanical embedding scheme all interactions between the two subsystems are handled at the force field level. Chemical bonds between QM and MM atoms are modeled by harmonic potentials, as well as the angles that are defined by one QM atom and two MM atoms, while dihedral torsions involving at most two QM atoms are commonly modeled by a periodic potential function. Non-bonded interactions (between atoms separated by three or more atoms) are also modeled by force field terms: the van der Waals and the electrostatic interactions.

In the simplest implementation of mechanical embedding scheme, the electronic wavefunction is evaluated for an isolated QM subsystem. Therefore, the MM environment cannot induce polarization in the QM region. For calculating the electrostatic interactions between the subsystems, one can either use a fixed set of charges for the QM region, or re-compute the partial charges on the QM atoms at every integration step if we are running a MD simulation. In geometry optimizations an update of charges at the end of each optimization is required to load QM charges for the next step.

Lennard-Jones parameters are normally not updated. Because of that, problems may arise if some changes occur in the chemical character of the atoms in the QM region, as changes in hybridization state of the atoms. However, since Lennard-Jones potential is a relatively short-ranged function, the error, introduced by keeping the same parameters throughout the calculations, is most likely not very large [88].

2.6.1.2 Electrostatic embedding scheme

To improve mechanical embedding, a new scheme that includes polarization effects in the boundary interactions was developed, the electrostatic embedding scheme. In this scheme, electrostatic interactions between the two subsystems are handled during the calculation of the electronic wavefunction. The MM atoms can polarize the electrons in the QM subsystem. However, the atomic charges of the MM atoms have been parameterized to provide the realistic description of an MM system, rather than a physically correct charge distribution. The main doubt is if this strategy is realistic or not. In reality, the interaction between the systems, is not only due to the electrostatics between charged atoms, but also due to the polarization, exchange, charge transfer, dispersion and Pauli repulsion. In force

fields, only the combination of atomic charges and Lennard-Jones parameters provides a reasonable description of all these effects taken together. In other words, not only the MM charges but also the Lennard-Jones parameters would need to be parameterized for use in the electrostatic embedding scheme.

A further problem is the risk of over polarization near the boundary. The point charges on the MM side of the interface may attract or repel the electrons too strongly, leading to electron density spilling out into the MM region. This over polarization could be serious if a large basis set, with polarization and diffuse functions, or plane waves are used in the QM calculations.

2.6.1.3 QM/MM boundary and link atoms

If the QM and MM subsystems are connected by chemical bonds, some care has to be taken. A straightforward cut through the QM/MM bond creates one or more unpaired electrons in the QM subsystem that, in reality, are paired with electrons belonging to the atom on the MM side.

In cases where there are non-bonding interactions between QM and MM atoms, there are no problems. In some other situations, it is necessary to split a molecule between QM and MM region, which means that there will be covalent bonds between the two regions. These bonds must be treated in a different way because the presence of unpaired electrons at the boundary of the QM region can dramatically change the electronic structure of the QM system. The most widely method used in application is the simple link atom method.

Link atoms are the easy solutions to treat the boundary between QM and MM regions. The strategy of using link atoms consists in introducing atoms at an appropriate position along the bond vector between the QM and MM atoms. Hydrogen is the most often atom used as link atom, but methyl groups and fluoride atoms can also be applied. The link atoms are present only in the QM calculation and they are invisible to the atoms in the MM region. Each link atom introduces three additional degrees of freedom to the system.

CHAPTER 3

COMPUTATIONAL METHODOLOGY

LINE OF WORK

3.1 Introductory note

As we have seen until now, the study of the catalytic mechanism of an enzymatic reaction is very important for future drug design. In the study of human renin, the knowledge of the catalytic mechanism with atomistic detail has an essential and determinant role in the development of new and more efficient antihypertensive drugs.

The starting point of any theoretical enzymatic study is a good tree-dimensional structure of the system. It is essential to identify the minimal system which mimics realistically the biological study system because a successful computational study greatly depends on the initial structure. The experimental data available on the literature are indispensable for a computational enzymatic study. These studies can offer structural information of enzymes and their substrates, obtained by protein crystallography and NMR spectroscopy techniques, which are crucial for starting this kind of studies.

The Protein Data Bank (PDB) [89] is the best known and the more complete archive of structural data of biological macromolecules with three-dimensional structures. Sometimes, when initiating a study there several crystallographic structures of the same protein. In these

cases, it is necessary to pick the “best” structure and the most usual selection criteria is the crystallographic resolution, which indicates how well defines the crystal structure is, the presence of ligands and how complete the structure is. The selection of the initial structure of a system must be made with special attention in order to guarantee the success of the study.

3.2 Preparation of the system

First of all, this work starts with a search in PDB database by *“human renin”*. After the analysis of all published structures, the initial structure was taken from PDB 2REN, which contains the crystallographic structure of free human recombinant renin at 2.5 Å resolution [42]. The selection was based on the resolution of the structure and how complete the structure is. With the structure of this enzyme three different models were built.

3.2.1 Ren:Ang$_{dodecapetide}$:W1 model

In the first model the substrate was modelled from the N-terminal peptide tail residues (Val3 – Asn14) from human angiotensinogen, from PDB 2X0B, at 4.4 Å resolution [90]. Despite the low resolution of this structure, part of it was used to modulate subtract because it was the only published crystal structure containing human angiotensinogen (Figure 7A)

The catalytic water molecule was modelled in the active site of the enzyme.

3.2.2 Ren:Ang$_{dodecapetide}$:W$_1$:W$_2$ model

The second model was similar to the previous one, but with an extra structural water molecule, establishing hydrogen bonds between the Tyr83 and Ser41 residues, as shown in Figure 7B. This water molecule has been described as important for the catalytic mechanism of aspartic proteases. With these two templates we will evaluate the role of it in the catalytic mechanism of human renin.

3.2.3 Ren:Ang$_{mutated}$ model

A third system was modelled using the optimized native structure. In this substrate, the Leu10 residue was replaced by a Phe residue (Figure 7C). This mutation was previously associated with preeclampsia, a common hypertensive disorder of pregnancy [91] .

The substrates were modulated in the renin cleft and then we align and superimposed the modeled system with the renin-angiotensinogen structure from PDB 2X0B.

The hydrogen atoms were added with Amber software X-leap [92] , considering the residues in their physiological protonation states. The exception was the catalytic Asp38 (proton donor residue) which was protonated according with previous experimental results and consistent with the proposed mechanism for aspartic proteases [4, 93]. Also with X-leap, eight counter-ions (Na^+) were added to neutralize the negative charge of the system. An explicit solvation model was used, with pre-equilibrated TIP3P water molecules in a truncated octahedral box with a minimum distance of 12 Å between the protein atoms and the box faces. The three systems had approximately 52650 atoms and the final models are similar to the system represented in Figure 7 and Figure 8.

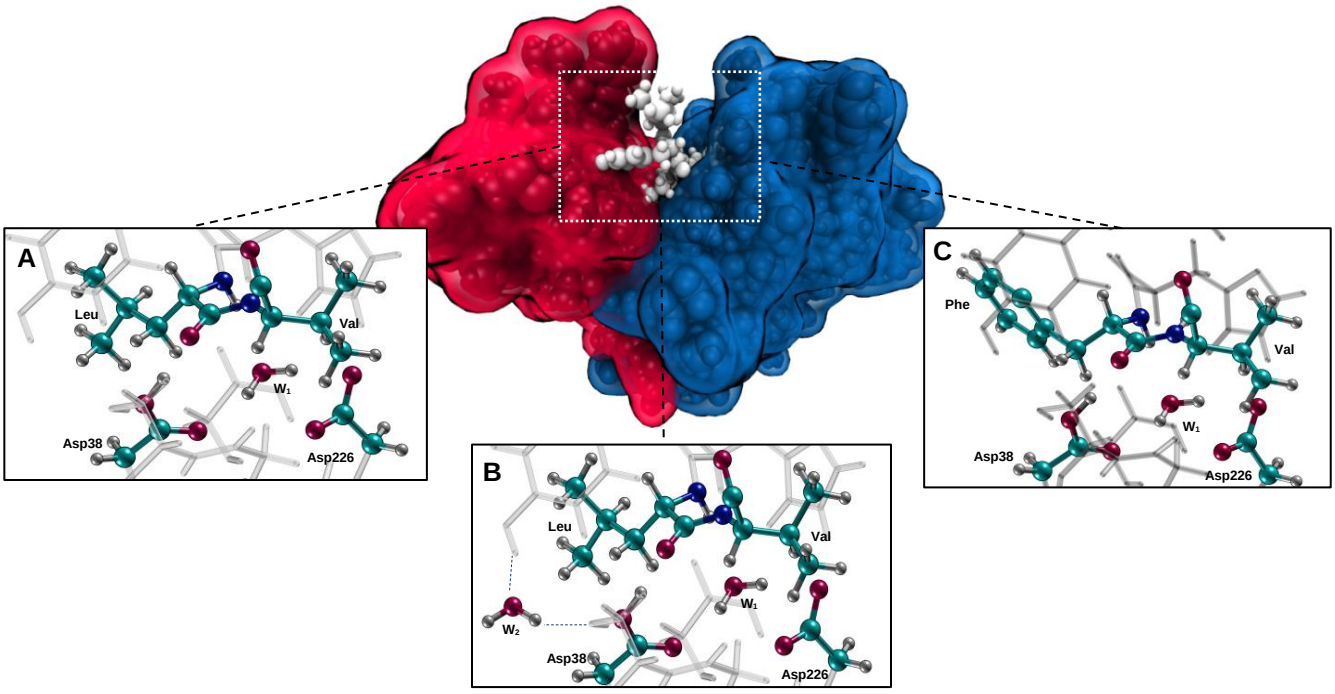


Figure 7. Representation of the different models used in the present work.

Catalytic Asp residues, Leu/Phe and Val residues of the scissile bond, and the catalytic (W_1) and structural (W_2) water molecules are represented in ball and sticks. (A) Active site representation of Ren:$Ang_{dodecapetide}$:W_1 model; (B) Active site representation of Ren:$Ang_{dodecapetide}$:W_1:W_2 model, where the second water molecule established H-bonds with Tyr83 and Ser41. (C) Active site representation of Ren:$Ang_{Mutated}$ model, where the Leu substrate residue was replaced by a Phe residue.

The geometry of the systems was minimized to release the bad contacts in the crystallography protein. The minimization was proceeding in two stages. In a first stage the protein is fixed and the position of the water molecules and counter ions was minimized. After, in a second stage, the full system was minimized.

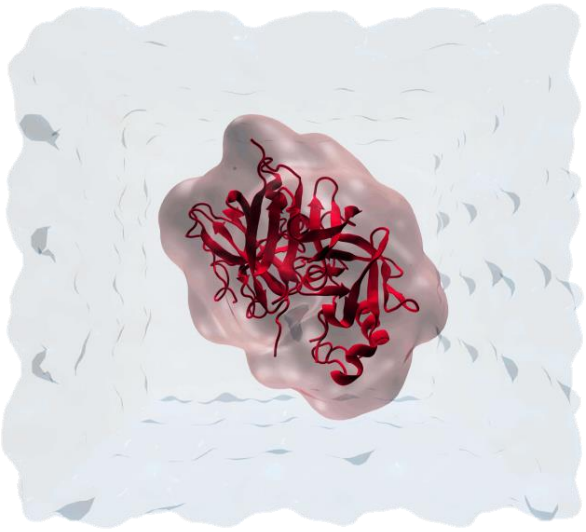


Figure 8. Representation of the system used in Molecular Dynamics Simulation.

The water molecules are represented in the box of 12Å around the protein.

3.3 Molecular dynamics simulations

In order to understand the behavior of the system in study, molecular dynamics simulations were performed. This computational method provides detailed information on the fluctuations and conformational changes of the system along time. The MD simulations were performed with Sander module implemented in AMBER09 software [94], using the Amber 2003 force field (parm03) [95].

After the minimizations a two-step MD simulation was run. First, an equilibration stage was performed using constant volume (NVT canonical ensemble) and considering periodic boundaries conditions. During this simulation the temperature was equilibrated using the Langevin temperature equilibration scheme, with an initial temperature of 0 K and final temperature of 310.15 K, during 200 ps. After, 6 ns of MD were run with constant pressure (NPT isobaric-isothermal ensemble) in the production stage. A collision frequency of 1 ps^{-1} was used to control the temperature. The SHAKE algorithm [96] was used in order to

consider the bond lengths involving hydrogen bonds and equations of motion were integrated with a 2 fs time-step. A 10 Å cutoff was used to truncate the no bonded interactions. The temperature was regulated with Langevin thermostat [97] to be maintained at 310.15 K during all the simulation. The results of the MD simulations were analyzed with PTRAJ module, which is included in the Amber package.

3.4 QM/MM calculations

3.4.1 Determination of the potential energy surface

The QM/MM calculations used to determine the potential energy surface (PES) along the catalytic pathway were run using the Gaussian 09 software [94]. The initial structures were obtained from the last minimization step before the MD simulations. In these models we have include the complex enzyme-substrate and the solvent water molecules in a radius of 5 Å around the complex. The total system was composed by 8204 atoms.

The ONIOM subtractive scheme [85], implemented in Gaussian 09, package was used to explore the PES energy along the catalytic reaction and we divided our system into two layers, a high level layer and a low level layer.

For the study of enzymatic reaction using QM/MM methods it is essential that the atoms included in either level are correctly selected. In the geometry optimizations, the high level layer included part of Leu10/Phe10 and Val11 residues of substrate, part of the lateral chains of catalytic Asp (Asp 38 and Asp 226) and the catalytic water molecule, with a total of 33 atoms Figure 9B.

The selection of the residues included in high level layer was based on previous experimental and theoretical studies, and specially in a previous study in the catalytic mechanism of mouse renin, which was developed by our group [4]. This high layer was treated with DFT with B3LYP functional and 6-31G(d) basis set [98, 99]. The remaining system was treating at MM level with parm99 force field (Amber). The hydrogen atoms were used as link atoms to simulate bonds and satisfy the valences of covalently bonded atoms across the QM/MM boundary. Residues with at least one atom within 20 Å around any QM atom were kept free, while the remaining part, beyond the 20 Å radius, were kept frozen. The described system was used in the initial optimization and the optimized structure

presented small differences when compared to the crystallographic prepared system. The same protocol was used for the three different models.

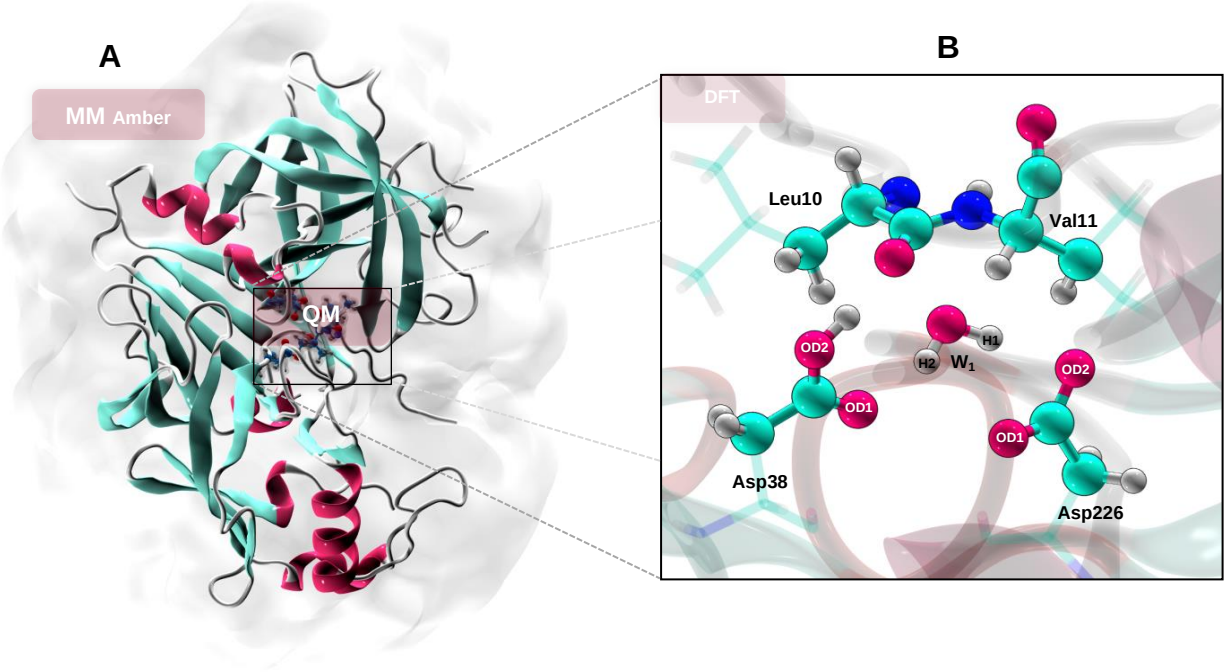


Figure 9. Representation of the system used in QM/MM calculations

(A) Representation of all system which includes the whole enzyme, the substrate (angiotensinogen dodecapeptide tail) and a shell of water molecules. The enzyme is represented with cartoon scheme. The system was divided in two layers: a high layer treated at QM level and a small layer treated with MM level. The high layer was represented with a large size in (B). The 33 atoms treated with QM level (DFT) are represented by balls and sticks.

A correct description of a TS structure is essential in understanding the progress from reactant to product as it is also important in developing of new drugs. Transition states were considering as the higher energy structures, which are described as good approximations to these states. The most reliable way of encountering a TS in a reaction profile involves the use of a linear transit scan. This consists in establishing the energy of several structures along a reaction coordinate, which will produce an energy curve where it is possible to identify the structure with highest energy. Therefore, we performed a linear transit scan along the reaction coordinate for each reaction step with increments of 0.10Å or 0.05Å to locate all the stationary points. To initiate the study of the chemical reaction the optimized structure was used as the initial structure.

The interaction between the two layers was described with mechanical embedding scheme, at the MM level. The atomic point charges for high layer atoms were recalculated in each step of the scan. These atomic point charges were obtained during a single point calculation using the Merz-Singh-Kollman scheme [100]. These calculations used the RESP (Restrained Electrostatic Potencial) charge fitting program [101, 102].

In the stationary points, single-point (SP) energy calculations were performed using the electrostatic embedding scheme and a more complex basis set (6-311++G(2d,2p)) [103].

3.4.2 Increase of the QM region

Afterwards, the high layer region (treated at QM level) was progressively increased (only for the Ren:$Ang_{dodecapetide}$:W_1 model) and new SP energy calculations, with B3LYP/6-31G(d) and B3LYP/6-311++G(2d,2p) levels, were performed. The increase of the QM model aimed to include the influence caused by residues around the active site, seeking to improve the energies of the catalytic reaction. This progressive increase was made in spherical shells, with a spherical radial increase centered on the catalytic water molecule. The new models are described below:

1) In the first new model, lateral chains of Leu10 and Val11 residues, from substrate, were added to the QM region. In Figure 10 the atoms that were added are coloured green. This system has 55 atoms in the high layer.

2) The second new model included, in QM region, the same residues present in previous models and also Ser38, and Tyr83 from enzyme and His9 from substrate. In the Figure 10 the new QM residues are coloured light blue. This system has 100 atoms in the high layer.

3) The third new model includes the same residues present in previous models and also Gly40, Ser84, Gly217 and Ala218. In Figure 10 the residues included in the new QM region are represented by the violet atoms. This system has 135 atoms in the high layer.

4) In the fourth new model Thr36, Val121, Leu213, Val214 and Thr216 residues and the complete chains of the catalytic Asp were added to the previous QM region. In Figure 10 the new QM residues are represented by the pink atoms. This system has 210 atoms in high the layer.

5) Finally, in the fifth and large model Leu75, part of Arg76, Gly80, Val122, Thrt79, Thr289, Ile330 residues were added to the QM region. In Figure 10 the new QM residues are coloured orange. This system has 322 atoms in the high layer.

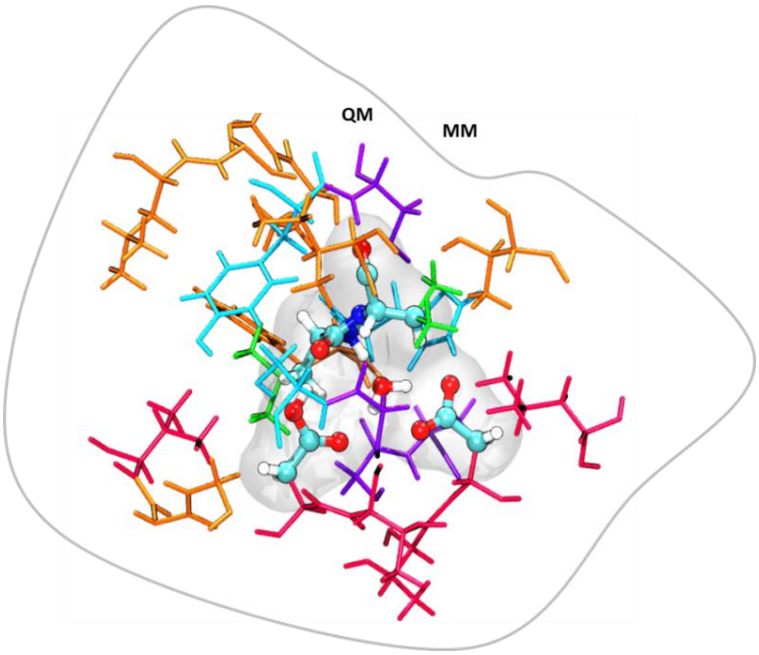


Figure 10. Representation of the atoms included in the different QM regions.

The different colors represent the residues that were added to the initial model. The initial QM region is represented by balls and sticks, and the atoms added to the first new QM region are represented in green. For the second model the light blue residues were added. Then the violet residues were also included in QM region. For the next model the pink residues were added and finally for the biggest model the orange residues were also included in the calculations.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introductory note

In the present work, we tried to understand the atomistic details of angiotensinogen hydrolysis catalysed by human renin. First, the results from MD simulations will be presented, in order to understand the dynamics of the system over time. Afterwards, the structures along the reaction mechanism and the hydrolysis energetic pathways will be discussed. Also in this chapter the role of a structural water molecule in the substrate cleavage and the difference between the renin reaction with wild type and mutated substrates will be evaluated. Finally, the human renin catalytic mechanism, which was studied in this work, will be compared to mouse one, which was recently published by our group [4].

4.2 MD simulation

As has been previously mentioned, in MD simulations three different systems were studied. The first one contains the angiotensinogen terminal dodecapetide, the second one a structural water molecule and angiotensinogen terminal dodecapeptide, and the last one

the Leu10Phe mutated substrate. After 200ps of equilibration, the simulations were run for 6ns and the results were analysed using PTRAJ tool, as previously mentioned.

4.2.1 Root Mean Square Deviation

The first analyzed parameter was the Root Mean Square Deviation (RMSd) that is a measure of the differences between the positions of specified sets of atoms. This measure is normally used to compare different conformations of the system under study during the time course of a MD simulation.

The RMSd values along the MD simulation performed on the first system are represented on Figure 11. It is possible to see that renin, represented by the blue line, does not change much its conformation during the simulation. After the equilibration simulation, renin does not vary much then 1Å. RMSd values for renin:dodecapeptide complex have large variations than the values obtained for the unbounded enzyme. At the end of MD simulation its position still oscillates about 2.5Å, however the complex seems to be stabilized.

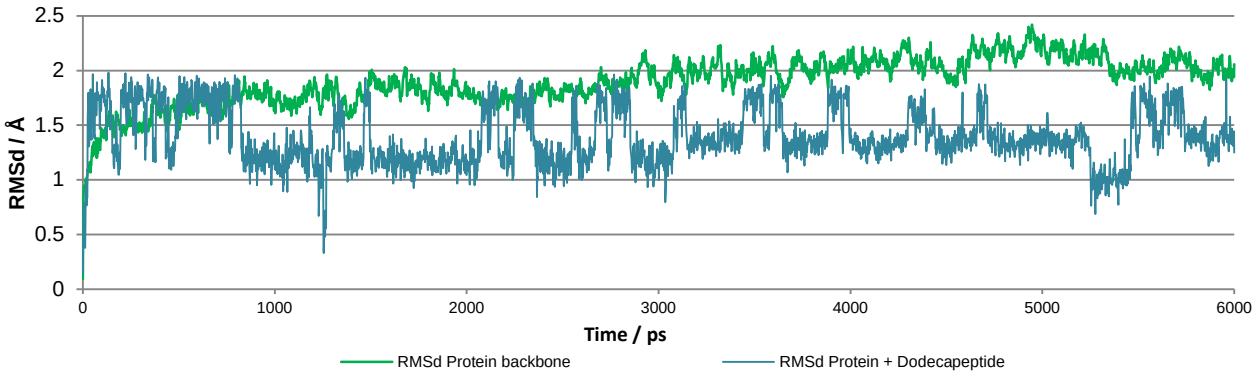


Figure 11. Renin and Renin:dodecapeptide RMSd along the MD simulation
Renin RMSd is represented in blue and Renin:dodecapetide RMSd is represented in green. Both values are stabilized at the end of MD simulation.

4.2.2 Root Mean Square Fluctuation

Then, the root mean square fluctuations (RMSf) was also analysed. The RMSf is a measure of deviation between the position of a particle (i.e. residue or atom) and a reference position (time-avaraged position).

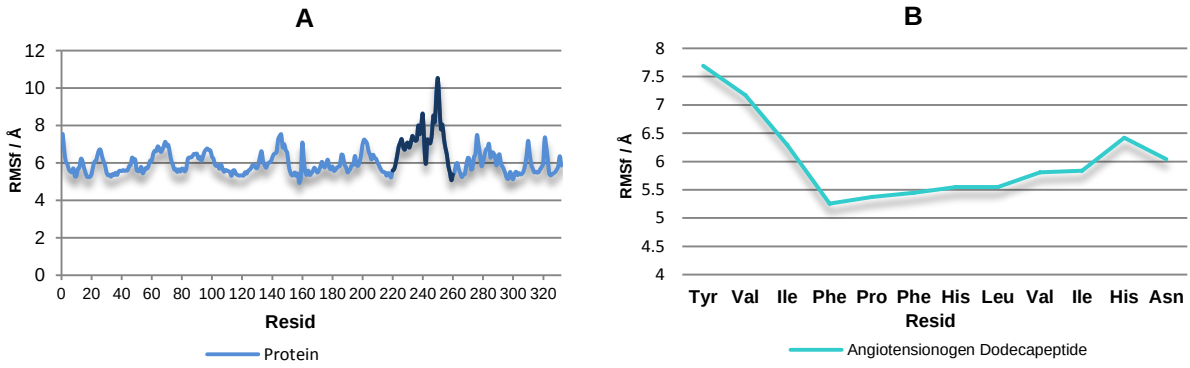


Figure 12. RMSf of Renin and Angiotensinogen residues along the MD simulation.

A) RMSf of renin resides. The dark blue line represents the protein region with large fluctuations. The Ang226 and resides around this catalytic resid are included in this region. B) RMSf of angiotensinogen dodecapeptide by residue.

The analysis of renin fluctuation (Figure 12A), during the MD simulation, shows that residues between the positions 220 and 260 (in complex pdb numeration) have the large fluctuation. These residues correspond to an external region of the protein and, probably because of it, they are more flexible. The residues that corresponds to the catalytic Asp38 and Asp226 (35 and 215 in complex pdb numeration) have also large fluctuations during the MD simulation.

The analysis of angiotensinogen fluctuation by atom shows that the first 2 residues (Tyr and Val) are the more flexible ones. They correspond to the first residues of angiotensinogen dodecapeptide tail. This residues have large and flexible lateral chains and because of it is expectable a large fluctuation.

4.2.3 Active site interactions

The length of interactions between catalytic atoms was also analysed during the MD simulation. This analysis is shown in Figure 13**.** The analysis of this figure shows some differences between the distances of represented atoms along the MD simulation. The major differences are visible after the 2ns of production, especially in the interactions between the water molecule and the catalytic residues.

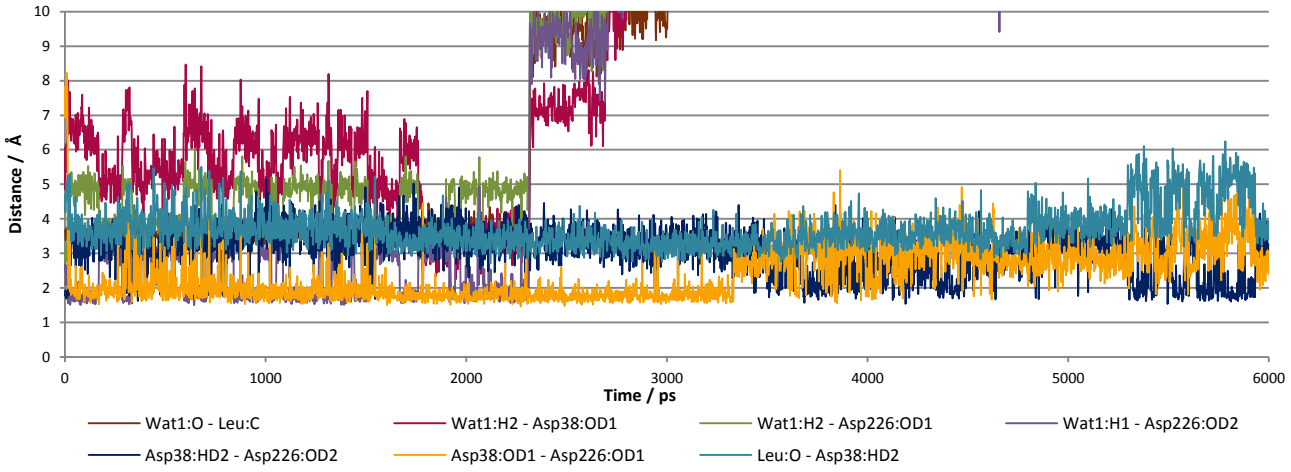


Figure 13. Representation of the main active site interactions throughout MD simulation.

Seven different interactions were represented. Brown line represents the interaction between catalytic water molecule and Leu residue of the scissile bond. Pink line represents the interaction between the catalytic water molecule and Asp38. Green line and violet line represent the interaction between the catalytic water molecule and oxygen atoms of Asp226. Yellow line represents the interaction between the Leu residue and Asp38. Finally the blue charts (light and dark) represents the interaction between Asp38 and Asp226. The distances for the other two systems were not represented because their behaviour was very similar to the presented one.

The distance between the water molecule atoms and some catalytic residues, at the end of the simulation, is approximately 50Å (not represented), which indicates that the interactions between these atoms completely cease after some time of simulation. During the course of MD simulation, it is also visible an approximation of the catalytic aspartates.

The visualization of the catalytic site residues and catalytic water molecule positions, during the dynamics, helps to understand the changes that happened throughout the simulation. Figure 14A shows the catalytic site of renin:angiotensinogen dodecapeptide complex before the MD simulation, and the Figure 14B represents the same system at the end of simulation. The starting structures for MD simulations were obtained by the X-ray crystallographic files of renin and angiotensinogen, followed by a superimposition with 2X0B structure, as mentioned before. At this point, catalytic Asp residues were in a coplanar position. The H-bond interaction between Asp35 proton and scissile bond carbonyl oxygen had a length of 1.74Å, and the catalytic water molecule was directed towards Asp226.

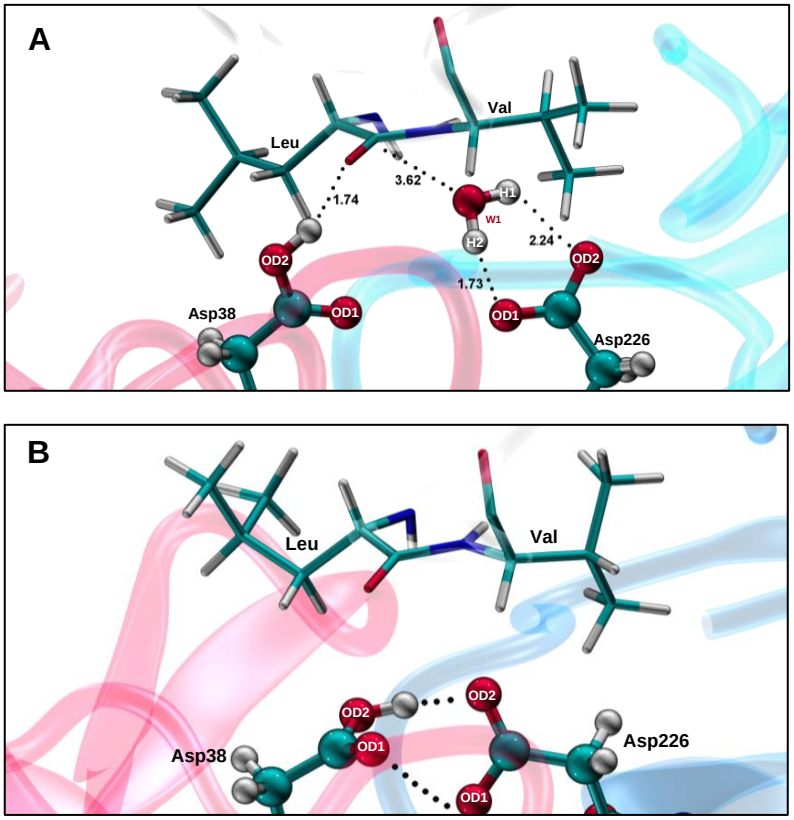


Figure 14. Representation of the renin:angiotensinogen system before and after the MD simulation.

(A) Before the MD simulation the catalytic Asp residues are in a planar position and they established H-bonds with the scissile bond carbonyl oxygen or with the catalytic water molecule. **(B)** After 2ns and until the MD simulation's end the water molecule moves away from the active center and the catalytic Asp residues remain directed to one another.

During the simulation, the water molecule gets apart from catalytic site and the Asp residues gets directed to each other. The observed geometries explain the changes in previous analyzed distances (Figure 13). This latter geometry is unfavorable to the occurrence of the catalytic reaction. The behavior of the other systems (with a second water molecule or with mutated substrate) was quite similar to the previously presented.

The catalytic water molecule was modeled on the active site of renin and, because of this, its position can vary more than the expectable, which may explain the diffusion of this molecule. This behavior has also been observed for other aspartic proteases.

The results obtained in the MD simulations allowed the analysis of system's conformational changes over the time. It was possible to observe that the system can change between two main positions, one conducive to the occurrence of the catalytic reaction and another unreactive. In this work, the MD simulations were performed only as an

initial study to know the behavior of the system along the time. The MD results have not been used for the catalytic mechanism study, however the structures obtained along the simulation could be useful for future studies on catalytic mechanism with different starting structures.

4.3 The catalytic mechanism of human renin

4.3.1 The Reactants structure

The model using the 2REN and 2X0B crystallographic structures, which was built in the beginning of this work (Ren:$Ang_{dodecapetide}$:W1 model) was optimized freely using to the default convergence criteria using the ONIOM method.

In the optimized structure, for the initial reactants, a water molecule (W1) interacts by hydrogen bonds with Asp38 (protonated) and Asp226 (unprotonated) residues of the catalytic site. The protonation of the catalytic aspartates is based on previous studies, as referred before. The active site of aspartic proteases is mono-protonated however it can be organized in two different forms. In the more stable form, the proton is allocated between the catalytic aspartate residues dyad, and it is shared in a hydrogen bond with both residues, which have a coplanar and symmetric conformation. However this structure is not as suitable for the catalysis as the second one. In this second form the Asp38 is protonated and it is H-bonded to the carbonyl oxygen of the substrate scissile bond [93]. In this work, we studied the second form and we used angiotensinogen tail (dodecapeptide) as substrate, which occupied the renin binding cleft.

In the reactants structure, the catalytic water molecule was placed between the catalytic aspartate dyad (Asp38 and Asp226) and the scissile bond of the substrate (Leu10 and Val11). The hydrogen bonds between these residues were represented in Figure 15A and in Table 2 and have the following values: H_{W1}-O_{Asp226} - 1.90Å and H_{W1}-O_{Asp38} - 2.07Å. At this stage, the Asp38 carboxylic group proton is orientated to the carbonyl oxygen of Leu10 and they establish a hydrogen bond with a length of 1.60 Å (Figure 15A). The Ser41 hydroxyl group forms a hydrogen bond with the carboxyl group of the Asp38, acquiring the latter one a more acidic character. It is also important to mention that two glycine residues (Gly39 and Gly220) also stabilized the aspartate dyad in the reactants structure, mainly the negative

charge of Asp226. There is also an important network of hydrogen bonds between Asp38-Ser41-Tyr80-Trp45 (data not shown).

4.3.2 The first reaction step

As we saw before, the first step of aspartic proteases proposed mechanism involves the formation of a gem-diol intermediate. Therefore, to study the renin catalytic mechanism, the reaction coordinate adopted for the first step was the distance between the catalytic water oxygen and the carbonyl carbon of the scissile bond (Figure 15A*).*

In this step, the protonated aspartate residue (Asp38) acts as an acid and the unprotonated one (Asp226) acts as a base. The enzyme activates the catalytic water molecule, which performs a nucleophile attack on the carbonyl carbon of the Leu10 residue from the substrate, concomitantly with the donation of a proton from Asp38 to the scissile bond carbonyl group. The optimized TS structure is shown in Figure 15B. With a detailed analysis on the TS for the first catalytic step (TS1), it is possible to see that the water molecule comes near to the carbonyl group of the scissile bond. The distance between O_{W1} and $C_{carbonyl}$ decrease from 2.77Å to 1.78Å. The proton of Asp38 comes closer to the carbonyl oxygen, but it is not yet transferred. The distance between these two atoms is 1.43Å at this point. The proton remains bonded to the Asp226 oxygen with a distance of 1.07 Å. This step does not cause great changes in the remainder system, which preserves its overall structure.

Then the structure of the first intermediate was found and, only at this stage, the proton from Asp38 has been completely transferred to the carbonyl carbon (bond length of 1.01Å). At the same time, during this step, one proton from the catalytic water molecule approaches to the carboxylate of negative Asp38 (bond length of 1.02Å) and a new covalent bond is established. This step leads to the formation of a gem-diol intermediate, in which two hydroxyl groups are attached to the carbon of the scissile bond and the catalytic Asp38 and Asp226 are negatively charged and neutral, respectively. At this step, the hybridization of the carbonyl carbon changes from sp^2 to sp^3, as expected, and the scissile bond carbon's geometry becomes tetrahedral instead of planar. In the intermediate structure the two hydroxyl groups are oriented to the Asp38 and they establish hydrogen interactions with both oxygen atoms of the carboxyl group with distances of 1.66Å and 1.64Å.

Although the similarities of the first intermediate with the gem-diol intermediates from other aspartic proteases, the TS1 from human renin is therefore different from the one proposed for other aspartic proteases like BACE-1 or HIV-1 proteases, in which a hydroxyl anion is formed at this stage [50, 93].

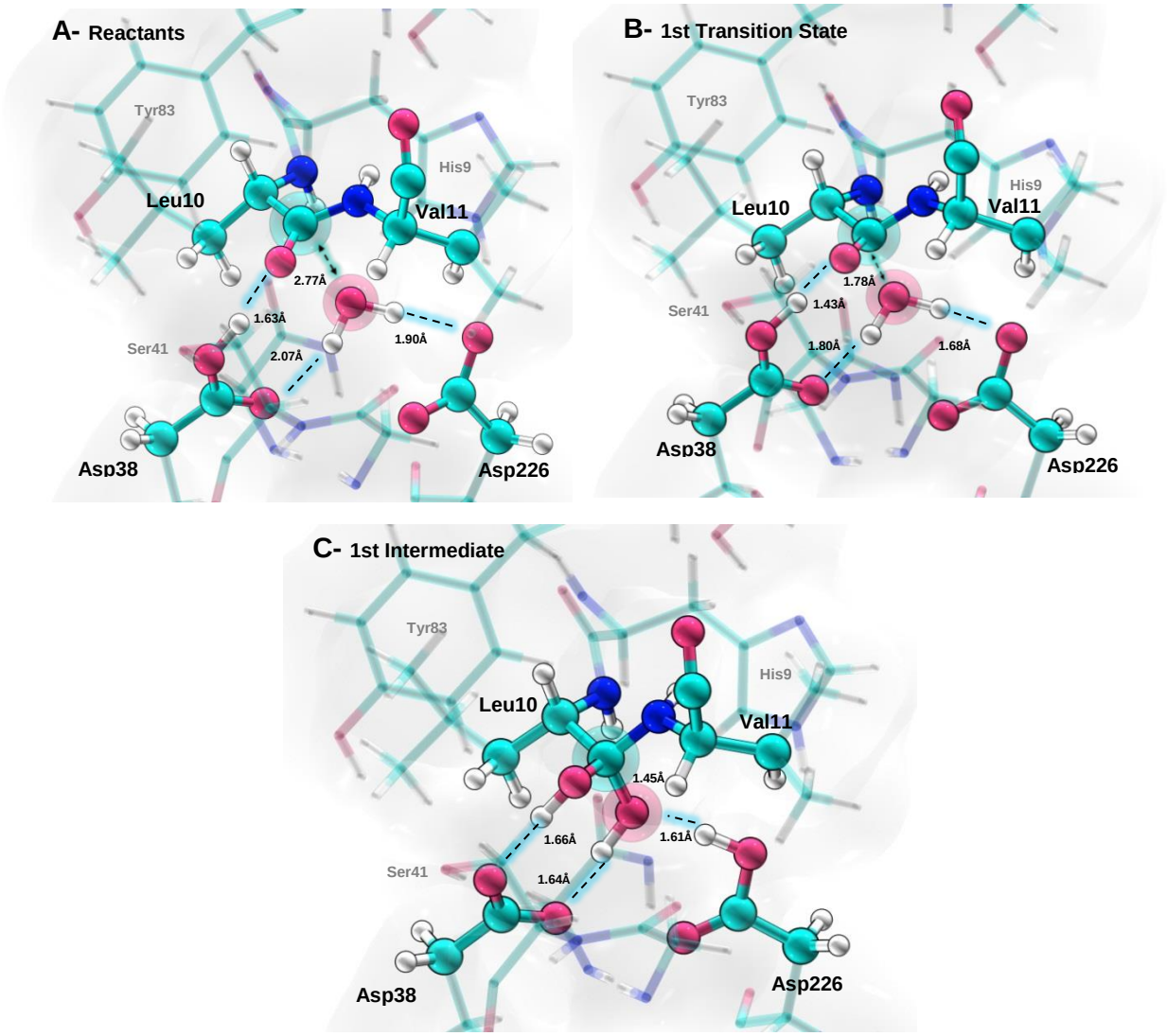


Figure 15. Reactants, TS1 and INT1 obtained geometries for Ren:Ang$_{dodecapeptide}$:W1 model.
A) Representation of the structure used as reactants. The adopted reaction coordinate are represented by the arrow between Leu10:C and W1:O. B) Representation of the TS1 structure. At this step the W1:O is near Leu10:C. C) Representation of the gem-diol INT1 structure. The main interactions between the atoms in the active site are represented in the figure. Other interactions are represented in**Table 2**.

It is also important to observe the behavior of the residues around the active site. In the TS1 and INT1 geometries the network of hydrogen bonds between Asp38-Ser41-Tyr80-

Trp45 is also kept as in the reactants (data not shown). The Ser41-Asp38 H-bond is shortened along the reaction coordinate from 1.85Å in reagents to 1.80Å and 1.68Å in TS1 and INT1 respectively. This residue stabilizes the carboxylate during the deprotonation of Asp38.

4.3.3 The second reaction step

Studies in other aspartic proteases have showed that the catalytic mechanism of this kind of enzymes occurs by two steps. After the formation of the gem-diol intermediate, the amide nitrogen is located in a better position to get the proton from Asp226, and this transference facilitates peptide bond cleavage. At same time the proton returns from the gem-diol intermediate to the Asp226. Because of this, we started our second step reaction from the INT1 structure, and the reaction coordinate adopted was the distance between the Asp226 proton and the nitrogen of the scissile bond. Figure 16 shows the second TS (TS2) and the second intermediate (INT2) geometries obtained in the second mechanistic step.

Comparing the INT1 and TS2 geometries, some important differences are evident. The approach of the Val11 nitrogen to Asp226, leads to the break of the H-bond between Asp216 and gem-diol hydroxyl and this Asp residue change its position with a rotation of its lateral chain. In the TS structure the proton from Asp226 is still not completely transferred to the amide nitrogen (1.30Å), but in the second intermediate the complete proton transference occurs (Figure 16B). Along this second mechanistic step, the substrate scissile bond stretches to 1.57Å at TS2 geometry and 1.6Å in INT2 geometry. This last distance means that this bond has not been completely cleaved at this step.

During this step, the proton transfer back from the gem-diol hydroxyl to Asp38 was also not observed. These results are different from previous studies on aspartic proteases in these two aspects. The hydrogen bond network mentioned before was conserved also in the TS2 and INT2 geometries. Furthermore, in the INT2 structure the catalytic aspartates have negative charge, the amide nitrogen has a positive charge and the products are not completely formed. These results are also different from previous studies on β-secretase that suggests a reorientation of the Ser41 and disrupt of the H-bonds established with Tyr83 which was not observed in our results for human renin [50, 104].

At the end of this second step, the products are not fully formed and therefore we performed a third reaction step that will be explained bellow.

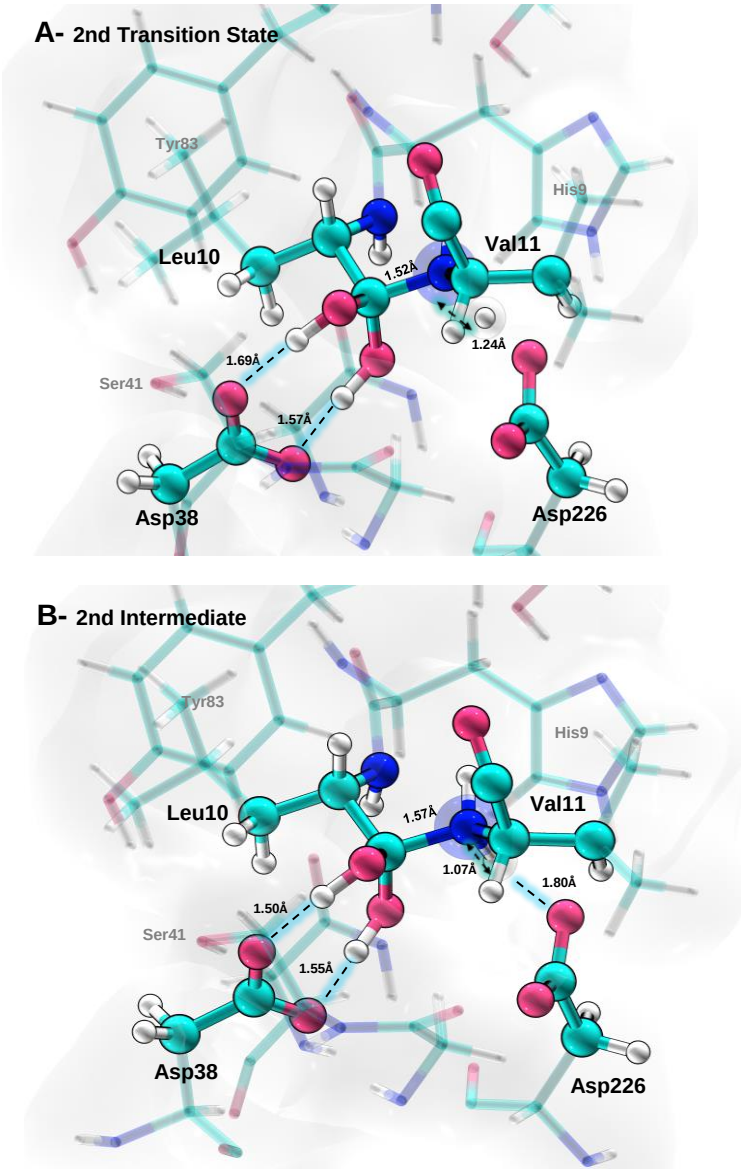


Figure 16. TS2 and INT2 geometries obtained for Ren:Ang$_{dodecapeptide}$:W1 model.

The reaction coordinate adopted is represented by the blue arrow between Asp226:H and Val11:N. A) Representation of the TS2 geometry. B) Representation of the INT2 geometry. The main interactions between the atoms in the active site are represented in the figure. Other interactions are represented in Table 2.

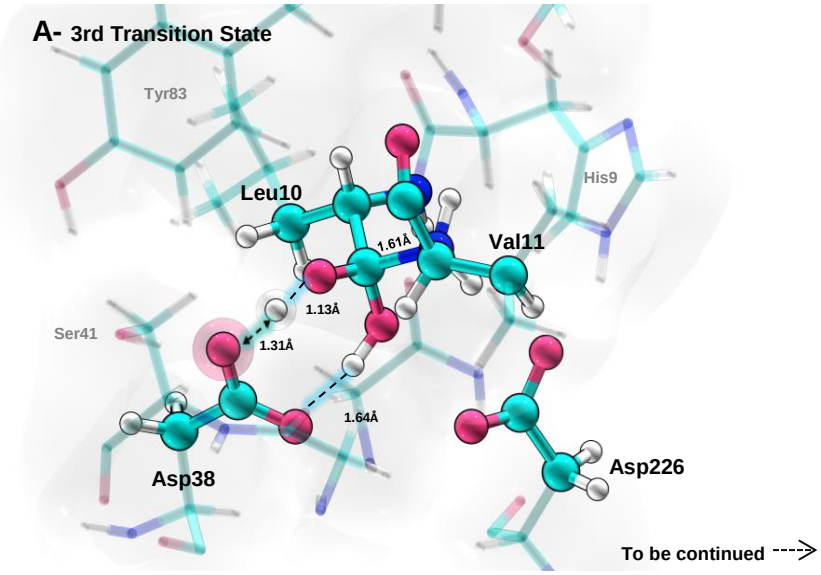
4.3.3 The third reaction step

We have performed a third mechanistic step in order to obtain the products of angiotensinogen hydrolysis by human renin. At this step the distance between the proton from one gem-diol hydroxyl group and the oxygen from Asp38 carboxyl group was used as reaction coordinate. The distances between carboxyl oxygen and the two protons from gem-

diol (carboxyl OD1:gem-diol$_{H1}$ and carboxyl OD2:gem-diol$_{H2}$) was tested as reaction coordinates, but only the first one leads to the scissile bond cleavage. During the scan the structure of TS3, followed by the structure of the reaction products (P) were obtained. During this step one gem-diol proton was progressively transferred to carbonyl oxygen of Asp38 and the scissile bond length increases and spontaneously breaks. The geometry of the final products shows a protonated catalytic dyad and two independent peptides, one with a charged carboxyl terminal group and another one with a neutral amine terminal (Figure 17B). Probably the initial protonation state (with protonate Asp226) will be obtained after the product release from the active center.

4.3.4 Summary of human renin reaction

The results obtained by our calculations shows that the reaction catalyzed by human renin is based on an acid/base mechanism and it occurs through three elementary reactions. First, the water molecule attacks the carbonyl group of the substrate leading to the formation of the gem-diol intermediate. The second step is initiated by the protonation of the peptide nitrogen of the scissile bond, and the third and last step is characterized by the transfer of the gem-diol proton back to the Asp38 accompanied by the complete cleavage of the substrate peptide bond and the formation of the products.



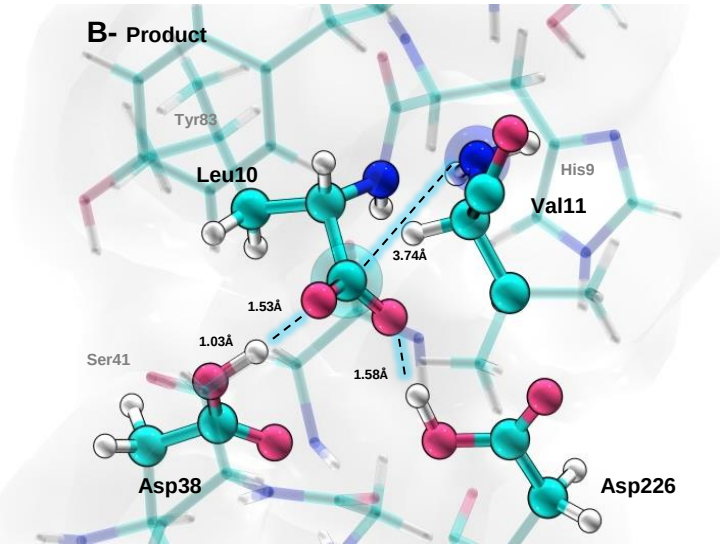


Figure 17. TS3 and Products geometries obtained for Ren:Ang$_{dodecapeptide}$:W1 model.

The reaction coordinate adopted is represented by the arrow structure between Asp38:OD1 and Gem-diol:H1. This step leads to the formation of two peptides. A) Representation of the TS3 geometry. B) Representation of the P geometry. The main interactions between the atoms in the active site are represented figure. Other interactions are represented in **Table 2.**

Table 2. Representation of the main interactions between the active residues and the residues around the active site during the reaction mechanism for Ren:Ang$_{dodecapeptide}$:W1 model.

Distances/Å	R	TS1	INT1	TS2	INT2	TS3	P
Active Residue							
Leu10:C-W1:O	2.77	1.78	1.45	-	-	-	-
Asp38:OD2-Leu10:O	1.63	1.43	1.01	-	-	-	-
Asp228:OD2-W1 H1	1.90	1.68	1.02	-	-	-	-
Leu10:C-Val11:N	-	-	-	1.52	1.57	1.61	3.44
Asp226:H$_{W1}$ -Val11:N	-	-	-	1.30	1.07	1.05	1.02
H$_{Asp38\text{-}gemdiol}$-Asp38	-	-	-	-	-	1.31	1.03
H$_{W1\text{-}gemdiol}$-Asp226	-	-	-	-	-	3.16	1.02
Residues around the active site							
Tyr83:H-Ser41:O	2.79	2.46	2.60	2.16	2.23	2.34	2.51
Ser41:H Asp38:OD2	1.85	1.80	1.68	1.72	1.73	1.75	1.80
His9:H Asp226:OD2	1.76	1.82	2.06	2.01	1.82	1.79	1.80

4.4 The role of a structural water molecule

Water molecules are an integral part of proteins, aiding in stabilization the protein fold and participating in their function.

Some studies showed that a structural water molecule (W2), located at the vicinity of the active groups, is conserved between aspartic proteases [54, 105]. These studies suggest that the W2 plays an essential role in formation of a network of hydrogen bounded residues (Asp38-Ser41-W2-Tyr83-Trp45) between the active carboxyl residues and the flap, which can assist the development of the catalytic reaction. In that way, in this work we compared the renin catalytic mechanism in the absence and presence of this water molecule. The mechanism with the W2 molecule will be described below.

4.4.1 The Reactants structure

Similar to the previously presented mechanism, in this new model, a catalytic water molecule, W1 molecule, also interacts with the Asp38 and Asp36 residues of the catalytic dyad. This W1 is also located between the catalytic dyad and the scissile bond of the substrate. The interaction between this water molecule and the residues around the active site is represented in Figure 18A. The distance between W1 and the oxygen of the carbonyl carbon, in the optimized structure, is slightly higher than that observed in the previous model however the remaining interactions are quite similar in both models. The protonation of the catalytic Asp residues was similar to the previous model, with Asp38 protonated and Asp226 unprotonated.

In this new model, the W2 molecule is located between Ser41 and Tyr83 residues. It established a network of interactions that starts in Asp38 to the Trp45 (not represented in Figure 18). All other residues, those are located around the active center, also maintain the same interactions observed in previous presented model.

4.4.2 The first reaction step

From the optimized structure described above, the reaction coordinate adopted was, as in the previous model, the distance between the oxygen from the W1 molecule and the carbonyl carbon of the scissile bond. During the scan along the reaction coordinate the geometry of the TS1 geometry was founded and it is represented in Figure 18B.

A detail analysis on this structure shows that the attack of the catalytic water molecule is also accompanied by simultaneous delivery of a proton from Asp38 to the oxygen of carbonyl group. The carbon atom of scissile bond and the oxygen atom from water molecule

are 1.75Å away from one another. Comparing this interaction with the same one obtained in the previous mechanism (1.78Å), they are quite similar. Regarding to the residues around the active site, they are also in very similar positions. The length of the hydrogen interaction established between the catalytic water molecule and the Asp38 was, at this point, 1.74Å, which is similar to the distance obtained in the previous mechanism (1.80Å). The interaction between this water molecule and the other catalytic Asp residue (Asp226) has a length of 1.67Å, which is also close to the value obtained earlier. The interaction between the Asp38 proton and the carbonyl oxygen was 1.39Å which is also similar to the 1.45Å observed in the previous mechanism. The slight differences between the geometry of the first TS for these two mechanisms are probably due to the presence of the structural water molecule.

Regarding the structure of the INT1 structure (Figure 18C), it was also very similar to the structure described for the previous mechanism. The gem-diol hydroxyl groups established hydrogen interactions with the carboxyl oxygen atoms of the Asp38 (1.65Å and 1.63Å compared to 1.66Å and 1.64Å from INT1 geometry obtained in the previous model). The interactions between the active site residues and around residues were also quite similar to those previously obtained. Therefore, we can conclude that in a structural view, the W2 molecule does not appear to affect the first step of the reaction catalyzed by human renin.

4.4.3 The second reaction step

As in the preceding mechanism, we start the second reaction step with the INT1 optimized structure, and the reaction coordinate adopted was the distance between the Asp226 proton and amide nitrogen of the scissile bond. The structure of the TS2 (Figure 19A) was found at a distance of 1.20Å between Asp proton and the amide nitrogen. A detail analysis of this structure shows that when we approach the proton to the amide nitrogen, the hydrogen interaction between the aspartate and gem-diol hydroxyl groups have been broken and the Asp226 change its position with a little rotation. The same behavior was observed in the previous mechanism.

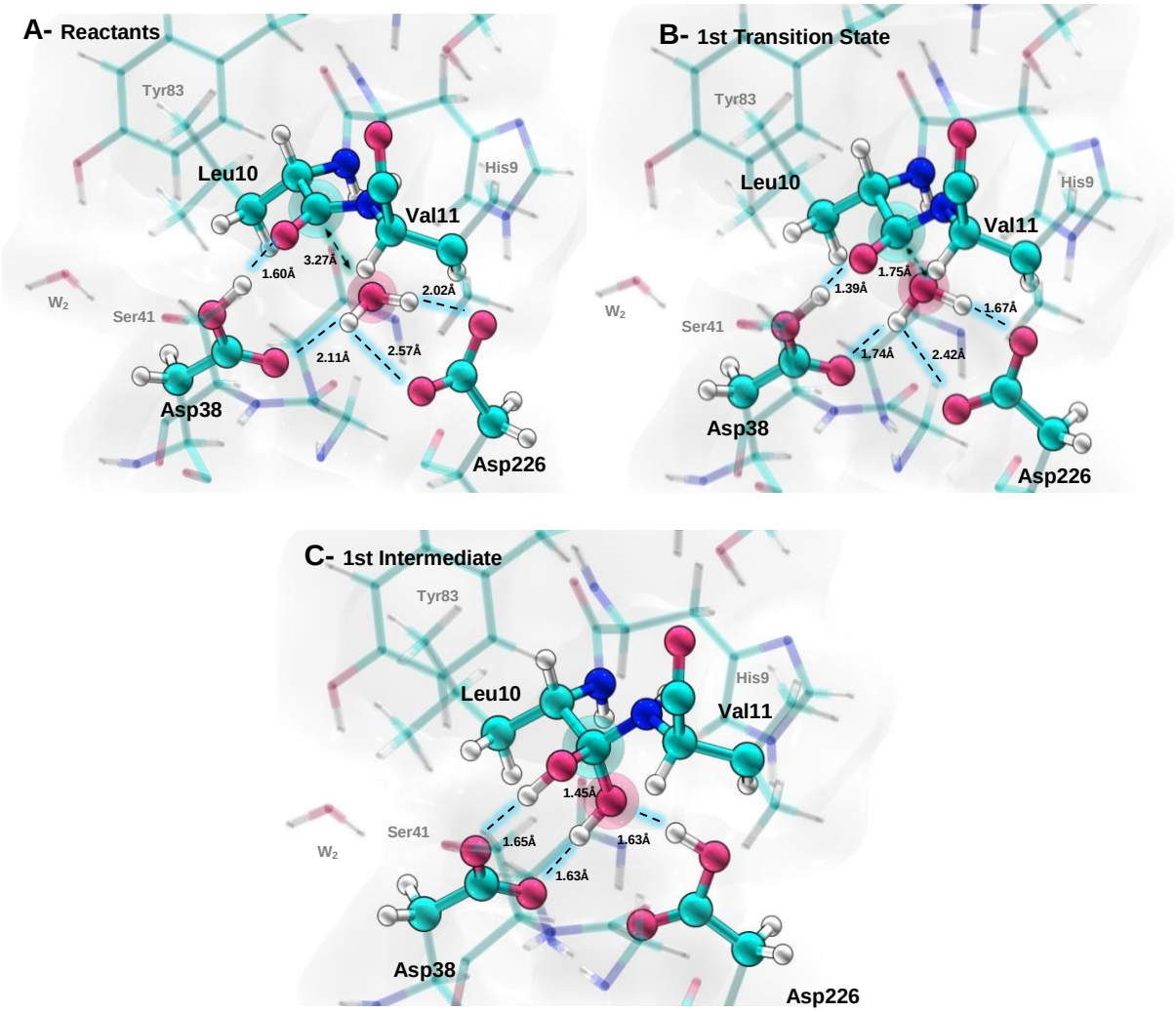


Figure 18. Reactants, TS1 and INT1 obtained geometries for Ren:Ang$_{dodecapeptide}$:W1:W2 model

A) Representation of the structure used as reactants. The adopted reaction coordinate are represented by the arrow between Leu10:C and W1:O. B) Representation of the TS1 geometry . At this step the W1:O is near Leu10:C. C) Representation of the gem-diol INT1 geometry. The main interactions between the atoms in the active site are represented in the figure. Other interactions are represented in Table 3.

Also in TS2 geometry, the scissile bond stretched to 1.54Å. All the interactions between the active site residues are maintained during this part of the second reaction step, and they are similar to the interactions observed to the previous mechanism. The Ser41 residue interacts with Asp38 through a 1.72Å hydrogen bond, as in the previous mechanism. With this result, we can conclude that the presence of the structural W2 is irrelevant for the active site rearrange during the TS2 formation.

The complete transfer of the Asp226 proton to the amide nitrogen only occurs in INT2, which is also similar to the previous presented mechanism without the structural water

molecule. Also as in the previous mechanism the scissile bond stretches to 1.57Å, which means, again, that this bond has not been completely cleaved at this point and the products are not completely formed. Therefore, a third mechanistic step was performed.

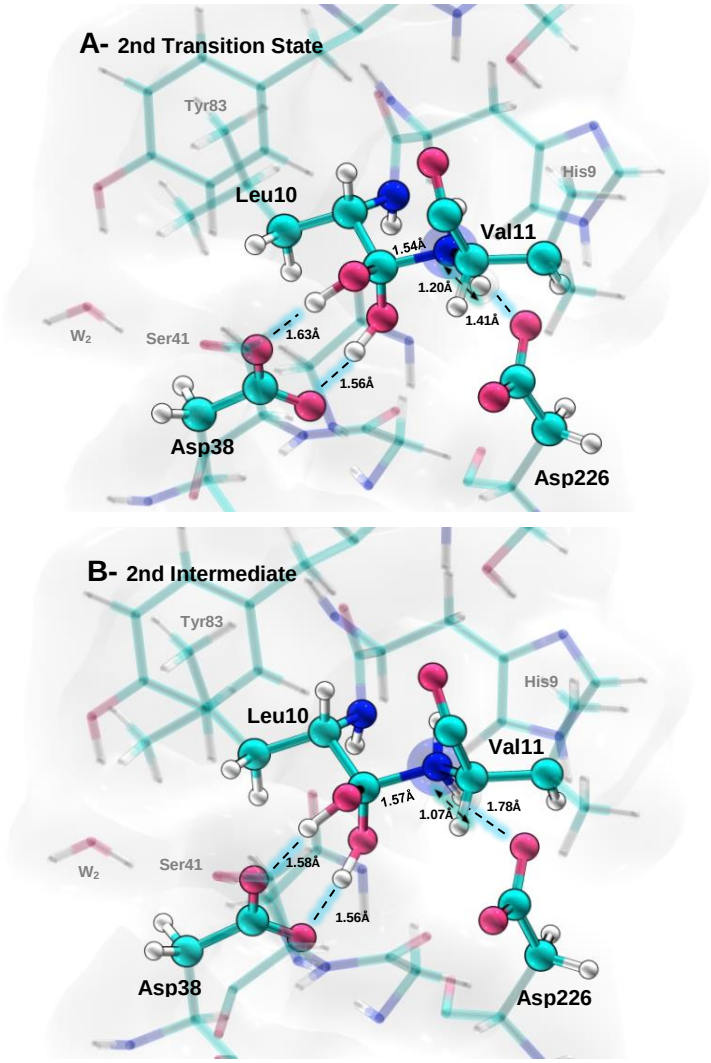


Figure 19. TS2 and INT2 geometries obtained for Ren:Ang$_{dodecapeptide}$:W1:W2 model.

The reaction coordinate adopted is represented by the arrow between Asp226:H and Val11:N. A) Representation of the TS2 geometry. B) Representation of the INT2 geometry. The main interactions between the atoms in the active site are represented in the figure. Other interactions are represented in Table 3.

4.4.4 The third reaction step

Again, likewise the previous mechanism, a third reaction step was achieved in order to obtain the reaction products. The adopted reaction coordinate is represented in the in Figure

20A. During the scan the TS3 and P geometries were obtained. The TS3 active site geometry retains its overall structure and all interactions are similar to the previous mechanism. As the reaction proceeded the hydroxyl gem-diol proton (H1) was progressively transferred to Asp38, and the scissile bond spontaneously broke. The rearrangement of the active site residues was also similar to the earlier mechanism.

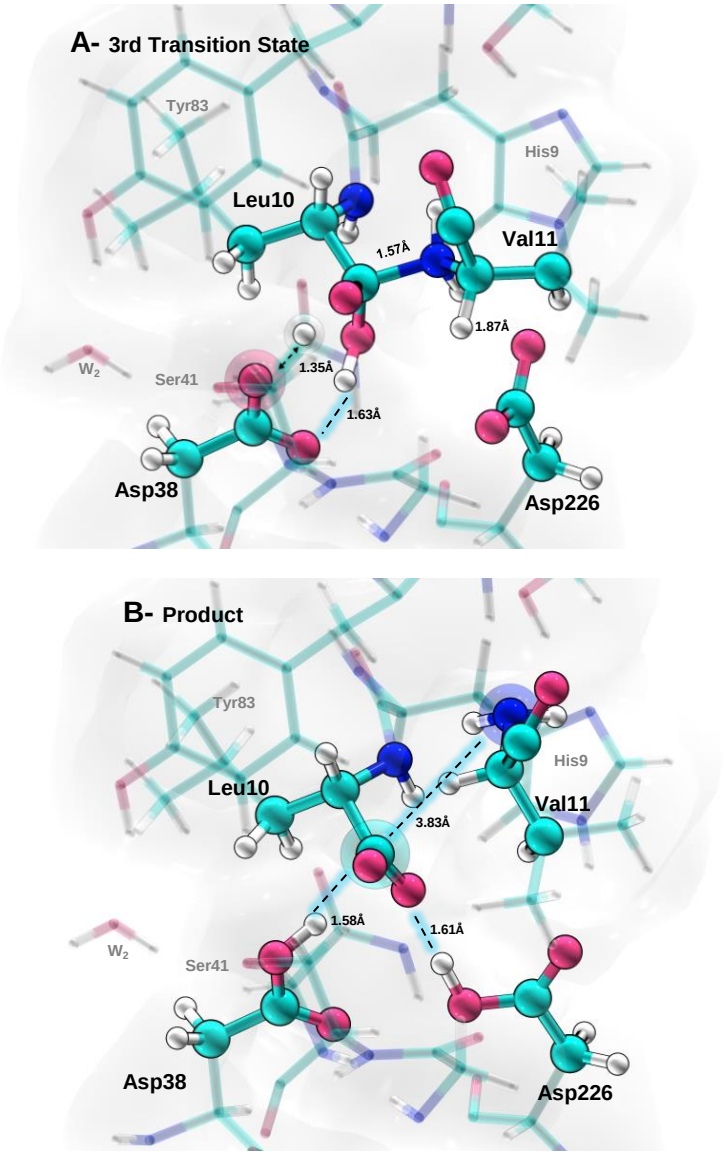


Figure 20. TS3 and Products geometries obtained for Ren:Ang$_{dodecapeptide}$:W1:W2 model.

The reaction coordinate adopted is represented by the arrow between Asp38:OD1 and gem-diol:H1. A) Representation of the TS3 geometry. B) Representation of the Product geometry. The main interactions between the atoms in the active site are represented in the figure. Other interactions are represented in Table 3.

In this QM/MM study, the influence of the structural water molecule on the catalytic mechanism of human renin has been investigated and its influence in the geometry of the active site residues seems to be insignificant. However, the present study does not measure the impact of the structural water molecule on the folding of the protein. Other studies suggest that this water is strategically placed to stabilize the conformation of the flap. The movement of the flap occurs in many aspartic proteases and this water molecule appears to be aiding in this mechanism. However the exact function of this water molecule is not fully understood, yet [45, 105].

Table 3. Representation of the main interactions between the active residues and the residues around the active site during the reaction mechanism for Ren:$Ang_{dodecapeptide}$:W1:W2 model.

Distances/Å	R	TS1	INT1	TS2	INT2	TS3	P
Active Residues							
Leu10:C-W1:O	3.23	1.75	1.45	-	-	-	-
Asp38:OD2-Leu10:O	1.60	1.39	1.01	-	-	-	-
Asp228:OD2-W1 H1	2.02	1.66	1.02	-	-	-	-
Leu10:C-Val11:N	-	-	-	1.54	1.57	1.59	3.80
Asp226:H_{W1} -Val11:N	-	-	-	1.20	1.07	1.06	1.04
$H_{Asp38\text{-}gemdiol}$-Asp32	-	-	-	-	-	1.35	1.02
$H_{W1\text{-}gemdiol}$-Asp215	-	-	-	-	-	1.63	1.01
Residues around the active site							
Tyr83:H-W2:O	1.82	1.80	1.82	1.78	1.78	1.78	1.89
W2:H-Ser41:O	1.86	1.91	1.87	1.84	1.85	1.87	1.94
Ser41:H Asp38:OD2	1.87	1.80	1.67	1.72	1.73	1.75	1.81
His9:H Asp226:OD2	1.74	1.82	2.05	1.95	1.82	1.81	1.85

4.5 Human renin catalytic mechanism with a mutated substrate

Preeclampsia is a common hypertensive disorder of pregnancy and it is associated with maternal, fetal and neonatal morbidity and mortality. Some studies have identified a mutation in angiotensinogen at the position 10 leading to the replacement of a Leu to a Phe, at the site of renin cleavage [91]. Due to the importance of this alteration, we also study the atomistic detail of the catalytic mechanism of human renin with mutated angiotensinogen.

4.5.1 The reactants structure

The structure of the system is very similar to the first presented model. The only difference between these two structures lies in the substrate. At the position 10, a Leu residue was replaced by a Phe residue. The geometry of the new reactants is represented in Figure 21A.

4.5.2 The first reaction step

Once again, the strategy for the study of the first mechanistic step was the reaction coordinate between the catalytic water oxygen and the carbonyl carbon of the scissile bond. During the scan, along the reaction coordinate, the structure of the first TS was identified and it is represented in Figure 21B.

Identical to the previous mechanisms, the water molecule nucleophile attack is accompanied by a synchronized delivery of a proton from Asp38 to the oxygen of carbonyl group. A detail analysis on TS1 geometry shows that the distance between the O_{W1} and $C_{carbonyl}$ decrease from 3.24Å in reactants structure to 1.70Å. About the residues around the active site, their structures are similar to the previous ones obtained in the other mechanism. In a general comparison between this TS geometry with the previous ones, it is possible to observe that with the mutated substrate, the distance between O_{W1} and $C_{carbonyl,}$ in TS1 structure is little smaller than the previous ones. The H-bond between Tyr83 and Ser41 and is also smaller than the preceding structures. Although this specific differences, the remainder system is identical to the reaction with wild type substrate. The smaller distances are observed probably because Phe residue is a larger residue than Leu residue.

The proton from Asp38 has been completely transferred to the carbonyl carbon at this first step like it happened in the reaction with wild type substrate. This first reaction step leads to the formation of a gem-diol intermediate, with a change in carbon hybridization from sp^2 to sp^3 and planar to tetrahedral respectively, as expected. The H-bonds between Asp38-Ser41-Tyr83-Trp45 is also preserved as in reactants and there are no relevant alterations in main geometry when it is compared to the reaction with the wild type mechanism.

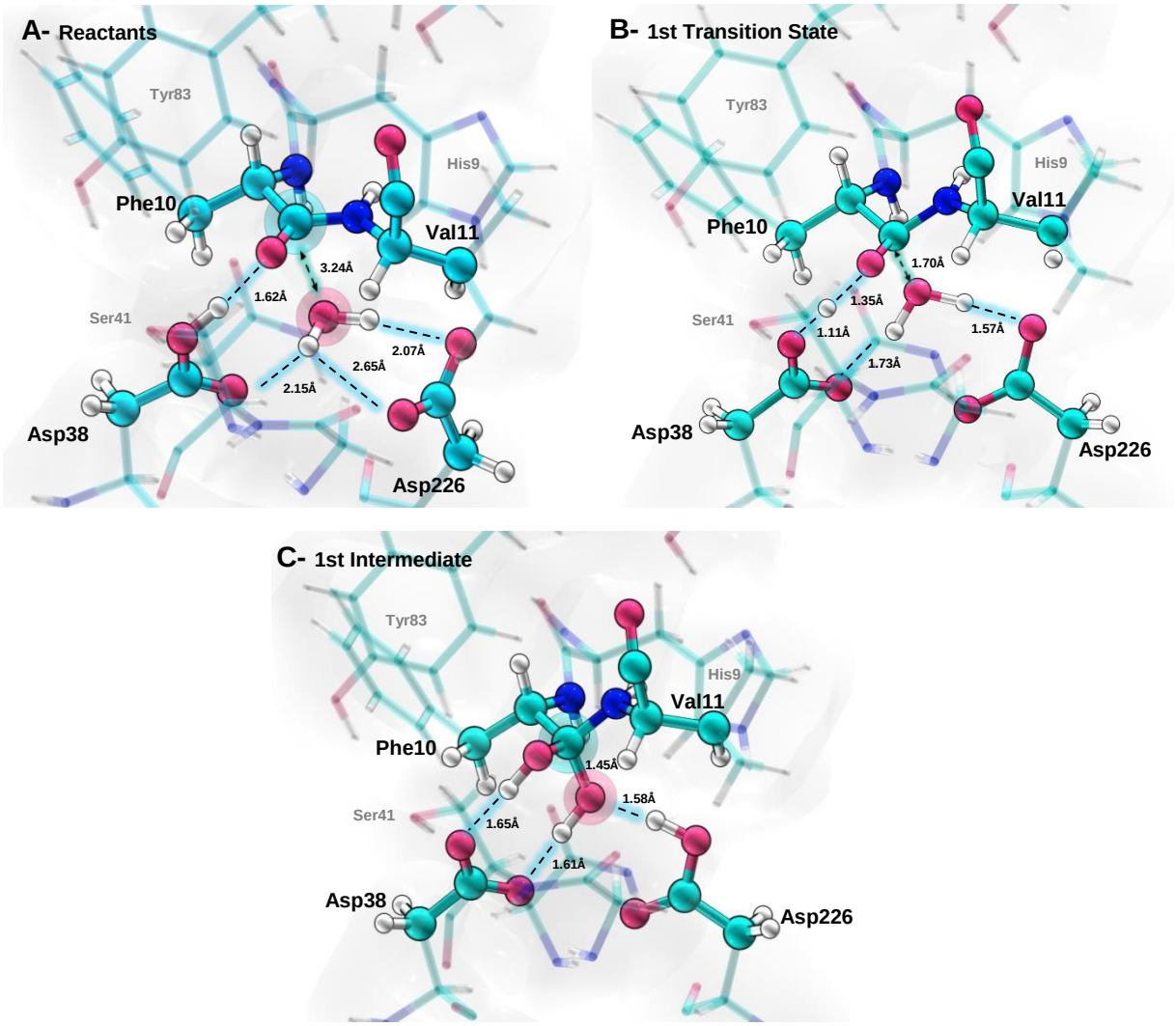


Figure 21. Reactants, TS1 and INT1 obtained geometries for Ren:Ang$_{Mutated}$:model.

A) Representation of the structure used as reactants. The adopted reaction coordinate is represented by the arrow between Leu10: C and W1:O. B) Representation of the TS1 geometry. At this step the W1:O is near Leu10:C. C) Representation of the gem-diol INT1 geometry. The main interactions between the atoms in the active site are represented in the figure. Other interactions are represented in Table 4.

4.5.3 The second reaction step

In the second mechanistic step the adopted reaction coordinate was the distance between the amide nitrogen and the Asp226 proton, as in the previous studies. In the TS2 geometry, the proton is located between this two residues and it only completely transferred in the INT2 structure.

In agreement with the reaction studied previously by us, and contrarily to other aspartic proteases, the scissile bond between de carbonyl carbon and amide nitrogen has not been cleaved yet and the proton from gem-diol has not been transferred back to the Asp38. In this structure, the catalytic Asp38 keeps their negative charge and the amide nitrogen gains a positive charge, in an analogous manner to the reaction with the wild type substrate. A third reaction step is also necessary for the cleavage of the mutated angiotensinogen.

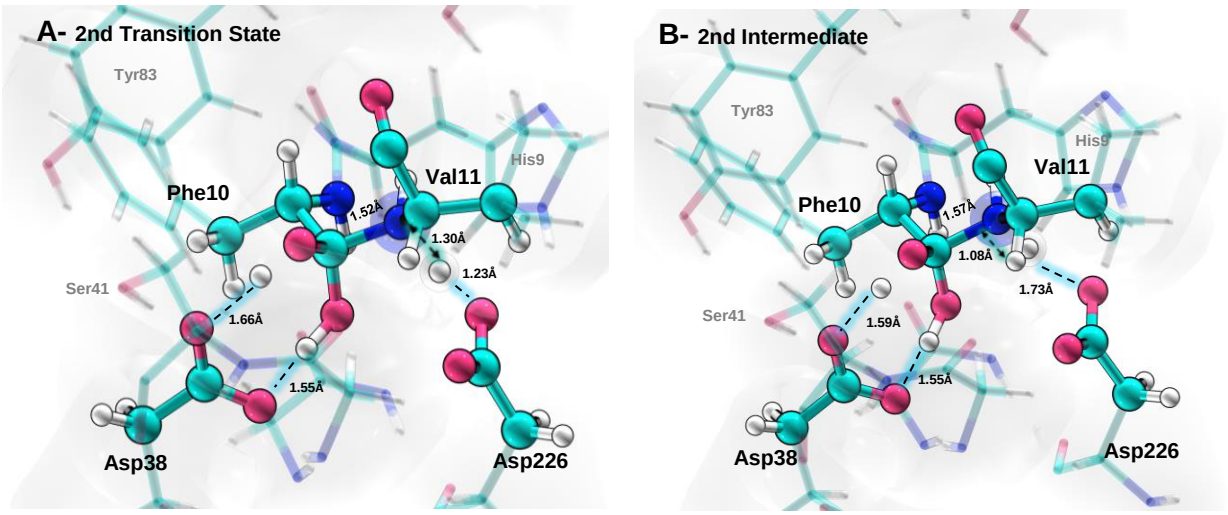


Figure 22. TS2 and INT2 geometries obtained for Ren:Ang$_{Mutated}$:W1 model

The reaction coordinate adopted is represented by the arrow between Asp226:H and Val11:N. A) Representation of the TS2 geometry. B) Representation of the INT2 geometry. The main interactions between the atoms in the active site are represented in the figure. Other interactions are represented in Table 4.

4.5.4 The third reaction step

In this new step, the distance between the proton from the gem-diol hydroxyl and Asp38 carboxyl oxygen, was used again as reaction coordinate. The geometry of the final products shows, again, a protonated catalytic dyad and two independent peptides, one with a charged carboxyl terminal group and another one with a neutral amine terminal, as it have been seen for the wild type substrate.

These results show that the catalysis of the mutated substrate is similar to the natural substrate. With these results, it is possible to conclude that the alteration of a Leu residue by a Phe residue, in angiotensinogen sequence, does not change the mechanistic reaction in a structural perspective. These results showed that the mutated substrate could be cleaved

likewise the wild type substrate and the presence of a different residues does not change the organization of the active site during the reaction.

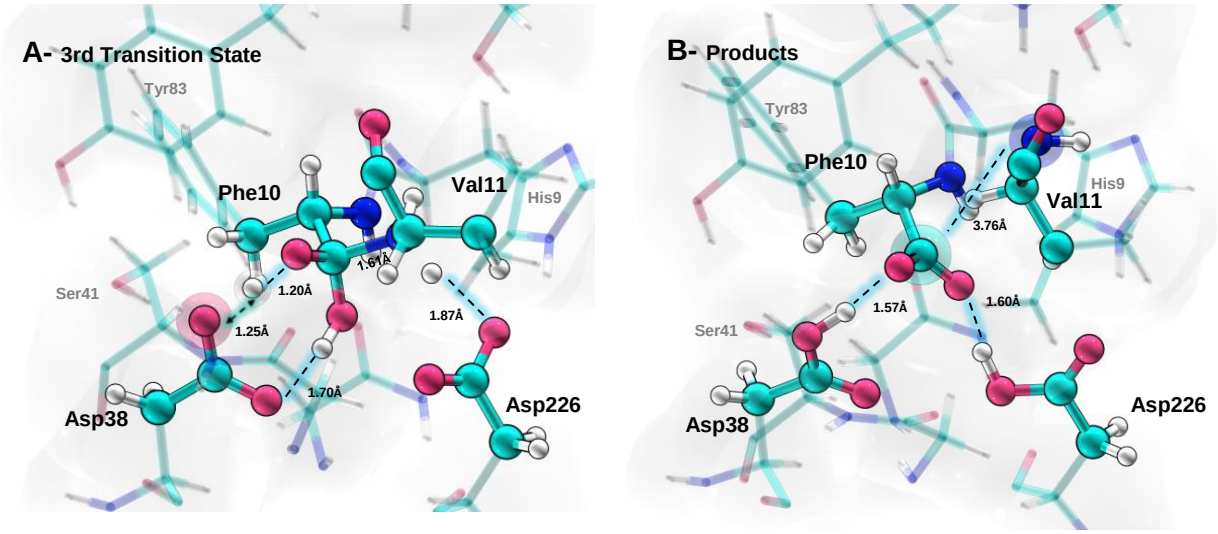


Figure 23. TS3 and Products geometries obtained for Ren:Ang$_{Mutated}$:W1 model.

The reaction coordinate adopted is represented by the blue arrow between Asp38:OD1 and Gem-diol:H1. A) Representation of the TS3 geometry. B) Representation of the Product geometry with formation of two peptides. The main interactions between the atoms in the active site are represented in the figure. Other interactions are represented in Table 4.

Table 4. Representation of the main interactions between the active residues and the residues around the active site during the reaction mechanism for Ren:Ang$_{dodecapeptide}$:mutated model

Distances/Å	R	TS1	INT1	TS2	INT2	TS3	P
Active Residues							
Phe10:C-W1:O	3.24	1.70	1.44	-	-	-	-
Asp38:OD2-Leu10:O	1.62	1.35	1.00	-	-	-	-
Asp228:OD2-W1 H1	2.07	1.57	1.00	-	-	-	-
Leu10:C-Val11:N	-	-	-	1.52	1.57	1.61	3.76
Asp226:H$_{W1}$ -Val11:N	-	-	-	1.30	1.08	1.06	1.04
H$_{Asp38\text{-}gemdiol}$-Asp32	-	-	-	-	-	1.25	1.02
H$_{W1\text{-}gemdiol}$-Asp215	-	-	-	-	-	2.98	1.01
Residues around the active site							
Tyr83:H-Ser41:O	2.48	2.18	2.27	2.09	2.12	2.20	2.05
Ser41:H Asp38:OD2	1.85	1.81	1.70	1.73	1.75	1.79	1.84
His9:H Asp226:OD2	1.75	1.86	2.09	2.06	1.86	1.81	1.86

4.6 Energies associated with human renin mechanistic pathway

The ONIOM energies were calculated along the reaction coordinate and the higher energies were associated with the Transition states geometries.

As we saw in the geometries description, according our calculations, the mechanism of human renin occurs through three reaction steps, while the mechanism of other aspartic protease (such as HIV-1 protease, β-secretase [50, 93]) occurs by two steps. The second step is characterized by the amide nitrogen protonation simultaneous with the protonation of deprotonated Asp residue and peptide bond cleavage, which was not observed in our results.

In this section, the energies associated with stationary points of human renin reaction will be presented and compared with available experimental data for this catalytic mechanism. Then the results of single point energy calculation, with large QM models, performed in order to include the influence of other residues in the energy associated with each reaction step, will be discussed.

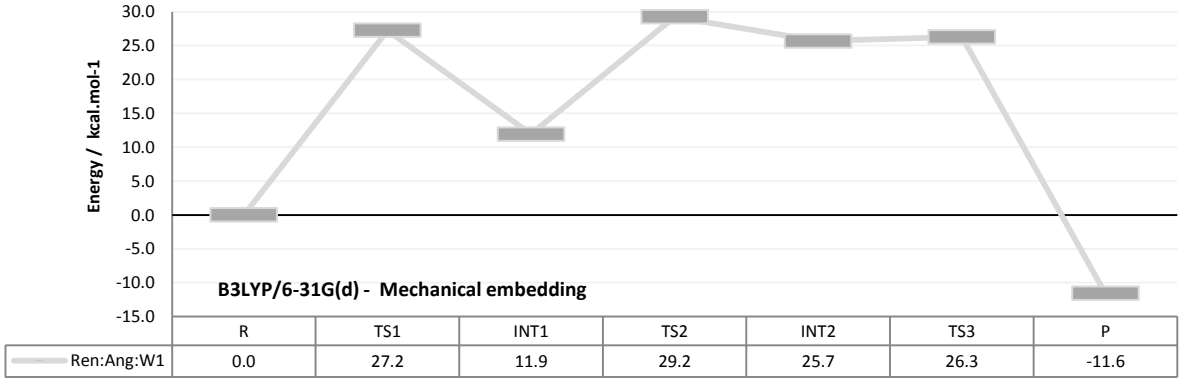
4.6.1 Energies along the reaction mechanism[2]

4.6.1.1 **Ren:Ang$_{dodecapetide}$:W1 model**

For the mechanism related to the first model, the calculated activation barrier for formation of the gem-diol intermediate, during the scan, was 27.2 kcal.mol^{-1}. This step was endergonic by 11.9 kcal.mol^{-1}. The second mechanistic step was characterized by activation energy of 29.2 kcal.mol^{-1} and it was also endergonic by 25.7 kcal.mol^{-1}. Finally, for the third mechanistic step, the calculated activation energy has a value of 26.3 kcal.mol^{-1}.Our calculations showed a global exergonic mechanism by -11.6 kcal.mol^{-1} (Figure 24).

[2] All energy values presented in this section correspond to relative energies (energy of reactants are defined as zero).

The experimental activation barrier obtained for the recombinant human renin, with recombinant human angiotensinogen, and with angiotensionogen terminal tetradecapeptide, obtained by previous studies, are 17.6 kcal.mol^{-1} (k_{cat} = 1.41s^{-1}) and 16.5 kcal.mol^{-1} (k_{cat} = 8.1s^{-1}), respectively [106]. The comparison of the experimental and computational data shows a significant difference between the described values. The energies were taken directly from the scan and the values represent enthalpy barriers instead of the activation Gibbs energies, which could partially explain the observed differences.



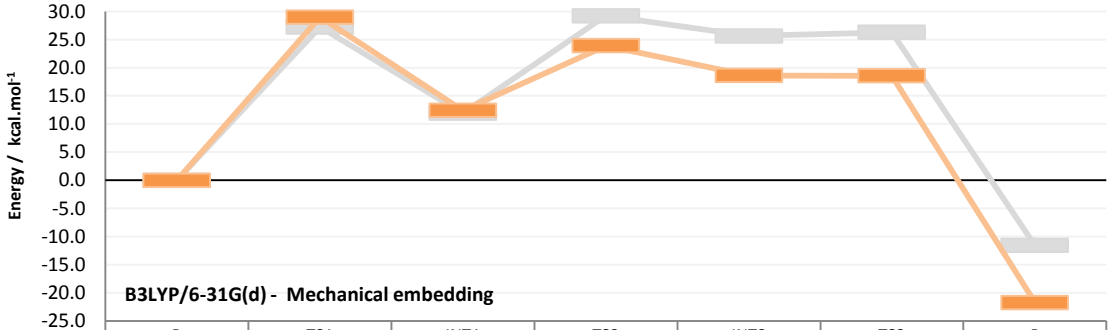
	R	TS1	INT1	TS2	INT2	TS3	P
Ren:Ang:W1	0.0	27.2	11.9	29.2	25.7	26.3	-11.6

Figure 24. Energetic pathway for angiotensinogen hydrolysis - Ren:Ang$_{dodecapeptide}$:W1 model
Barriers for this reaction were calculated at B3LYP/6-31G(d)//AMBER level and the energies are represented in kcal.mol^{-1}.

4.6.1.2 Ren:Ang$_{dodecapetide}$:W1:W2 model

For the second model mechanism the formation of a gem-diol intermediate was associated to 29.0 kcal.mol^{-1} and this first step was endergonic by 12.1 kcal.mol^{-1}. The formation of the second intermediate is associated with a value of 23.9 kcal.mol^{-1} and the second mechanism was endergonic by 18.6 kcal.mol^{-1}. The activation barrier for the formation of the third TS was 17.9 kcal.mol^{-1} and overall reaction is exergonic by -21.7 kcal.mol^{-1} (Figure 25). It is important to compare the energies obtained for the two models. The presence of structural W2 does not seem to have a big influence in the reduction of the relative reaction energies. The first reaction step had a larger relative energetic value in the presence of the W2. However, the second transition state is stabilized by the presence of this molecule, because it was associated with a low barrier. It is also important to refer that with the presence of W2 the catalytic reaction leads to more stable products. However,

when we compare these energies with the experimental ones, the activation barrier continues very high. These results show that the presence of the W2 is not essential for the occurrence of the angiotensinogen hydrolysis. In that way, the role of this molecule remains unknown.



	R	TS1	INT1	TS2	INT2	TS3	P
Ren:Ang:W1	0.0	27.2	11.9	29.2	25.7	26.3	-11.6
Ren:Ang:W1:W2	0.0	29.0	12.4	23.9	18.6	18.6	-21.7

Figure 25. Energetic pathway for angiotensinogen hydrolysis – Ren:Ang$_{dodecapeptide}$:W1:W2 model

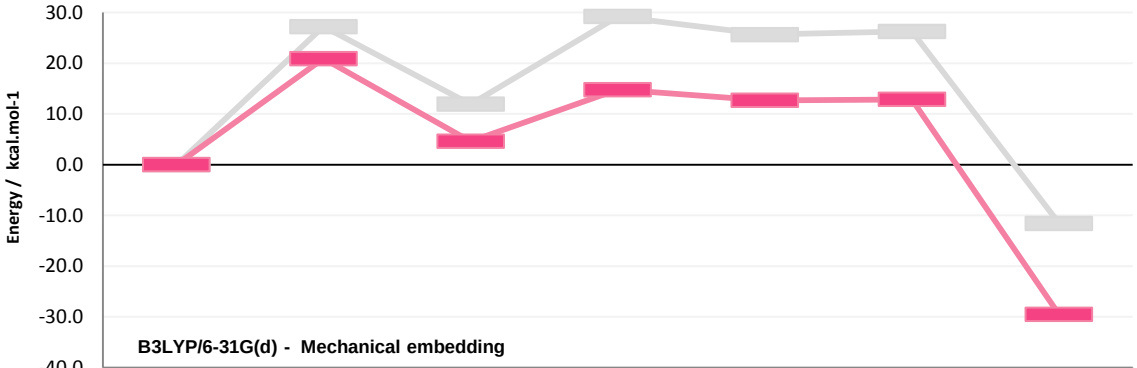
Barriers for this reaction were calculated at B3LYP/6-31G(d)//AMBER level and the energies are represented in kcal.mol^{-1}.Orange line corresponds to the energetic pathway for Ren:Ang$_{dodecapeptide}$:W1:W2 and grey line corresponds to the energetic pathway for Ren:Ang$_{dodecapeptide}$:W1.

4.6.1.3 Ren:Ang$_{Mutated}$ model

In the third model the activation barrier associated with the formation of gem-diol intermediate was 20.9 kcal.mol^{-1}. This first step was endergonic by 4.6 kcal.mol^{-1}. The second step had an activation barrier of 14.8 kcal.mol^{-1} and it is also endergonic with 12.7 kcal.mol^{-1} as associated energy. The activation barrier for the third TS was 12.9 kcal.mol^{-1}. This mechanism is also exergonic and leads to a very stable product (-29.5 kcal.mol^{-1}) (Figure 26).

Previous studies have shown an invariance of the cleavage site in a renin reaction with Leu10Phe mutated substrate. Furthermore they showed that the catalytic efficiency of renin reaction, measured as the ratio $k_{cat}/K_{m,}$ was higher for mutant substrate than for wild-type substrate [91]. The experimental activation barrier for this reaction, with the angiotensinogen terminal tetradecapeptide, is 18.3 kcal.mol^{-1}. This value is similar to the energy obtained in

this work. However, the energies obtained for the mutated substrate mechanism should be higher than the energies obtained for the wild type mechanism and this did not happened.



	R	TS1	INT1	TS2	INT2	TS3	P
Ren:Ang:W1	0.0	27.2	11.9	29.2	25.7	26.3	-11.6
Ren:AngMut	0.0	20.9	4.6	14.8	12.7	12.9	-29.5

Figure 26. Energetic pathway for angiotensinogen hydrolysis - Ren:Ang$_{Mutated}$ model

Barriers for this reaction were calculated at B3LYP/6-31G(d)//AMBER level and the energies are represented in kcal.mol^{-1}.Pink line corresponds to the energetic pathway for Ren:Ang$_{Mutated}$ and grey line corresponds to the energetic pathway for Ren:Ang$_{dodecapeptide}$:W1.

4.6.2 Single Point energy calculations

4.6.2.1 Single point calculations using electrostatic embedding scheme

As the first two presented mechanisms have been quite similar, the results for the first model was used to performed single point calculations to recalculate the energies in the stationary points, using the electrostatic embedding scheme. First, it was used B3LYP/6-31G(d) level to treat the QM region and electrostatic embedding scheme was used to treat the QM/MM boundary interactions.

The activation barrier for the formation of the first TS was 17.1 kcal.mol^{-1} (a decrease of about 10 kcal.mol^{-1} in comparison with the energy obtained with mechanical embedding). This new calculation indicates that the first reaction step is endergonic only by 2.2 kcal.mol^{-1}. The second mechanistic step are characterized by and activation barrier of 9.7 kcal.mol^{-1} and is also endergonic by 6.1 kcal.mol^{-1}. The third mechanistic step was characterized by

6.9 kcal.mol^{-1} and the relative energy associated to the products was -18.3 kcal.mol^{-1} (Figure 27). As said before, in mechanical embedding scheme the interactions between the two subsystems are included in the force field level and the MM environment cannot induce polarization in the QM region. On the other hand, in the electrostatic embedding scheme the electrostatic interactions between the two subsystems are handled at the DFT level and their influence in the polarization of the QM subsystem is included. In that way, with the obtained results, we can conclude that the influence of the MM residues in the QM region should be very important in the human renin catalytic mechanism.

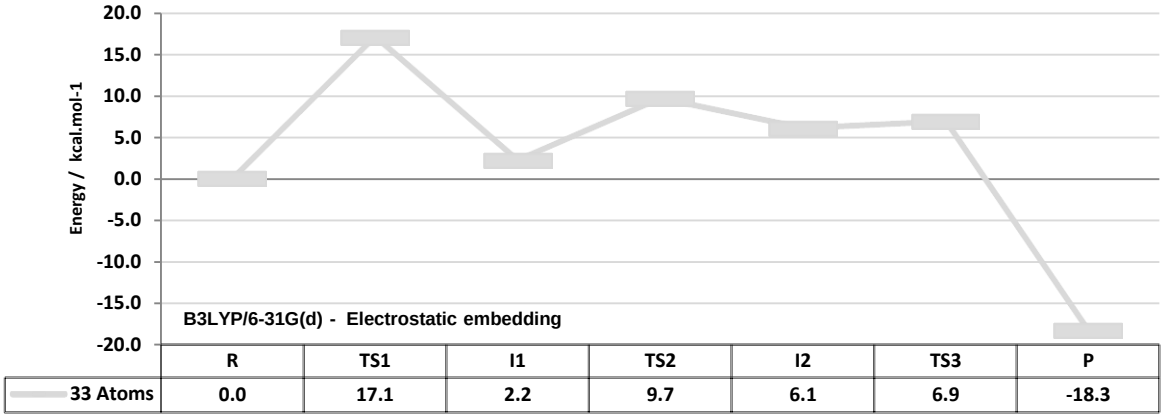


Figure 27. Single point calculation with B3LYP 6-31G(d) level and electrostatic embedding scheme.
Barriers for this reaction were calculated at B3LYP/6-31G(d)//AMBER level and the energies are represented in kcal.mol^{-1}.

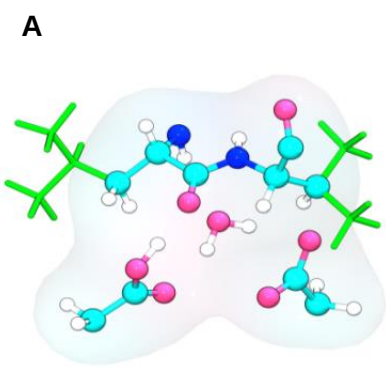
4.6.2.2 Single point calculation using large QM regions

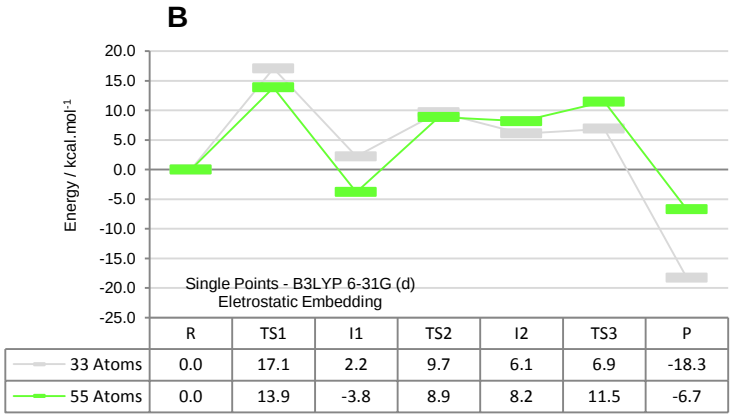
As mentioned before, in order to include the influence caused by residues around the active site in the energetic calculations, the QM region was expanded in a progressive way (spherical shells centered in the catalytic water molecule) and the energies were recalculated.

In the first model, only Leu10 and Val10 remainder lateral chains was added to stationary points and the QM region was increase from 33 atoms to 55 atoms.

The energy was recalculated at B3LYP/6-31G(d) level with electrostatic embedding scheme. The calculated activation barrier for the first mechanistic step formation was 13.9 kcal.mol^{-1} and contrarily to the previous calculations, the first reaction step was characterized by -3.8 kcal.mol^{-1} relative energy (exergonic). The calculated activation barrier for the formation of the second TS was 8.9 kcal.mol^{-1} and this step was endergonic by 8.2

kcal.mol^{-1}. For the third reaction step, the calculated activation barrier was 11.5 kcal.mol^{-1} and the formation of the products is associated with -6.7 kcal.mol^{-1} relative energy (Figure 28).

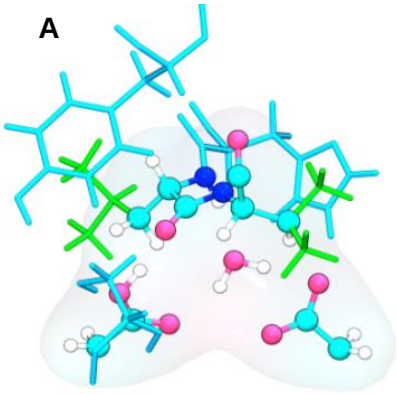


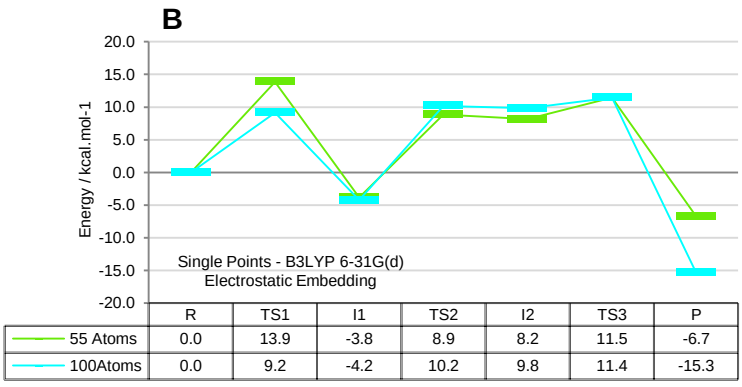


	R	TS1	I1	TS2	I2	TS3	P
33 Atoms	0.0	17.1	2.2	9.7	6.1	6.9	-18.3
55 Atoms	0.0	13.9	-3.8	8.9	8.2	11.5	-6.7

Figure 28. Single Point QM/MM calculations for a model with 55 atoms in high layer.
A) Representation of the high layer residues. B) Energetic pathway along the reaction (green representation).

In a second model His9, Ser38 and Tyr83 was added to the previous model. This system has 100 atoms in the high layer. The calculated activation barrier for the first TS was 9.2 kcal.mol^{-1} and this step was exergonic by -4.2 kcal.mol^{-1}.The second TS had an activation barrier of 10.2 kcal.mol^{-1} and the reaction step is endergonic by 9.8 kcal.mol^{-1}. The calculated barrier for third TS was 11.4 kcal.mol^{-1} and product formation is associated with -15.3 kcal.mol^{-1} relative energy (Figure 29).



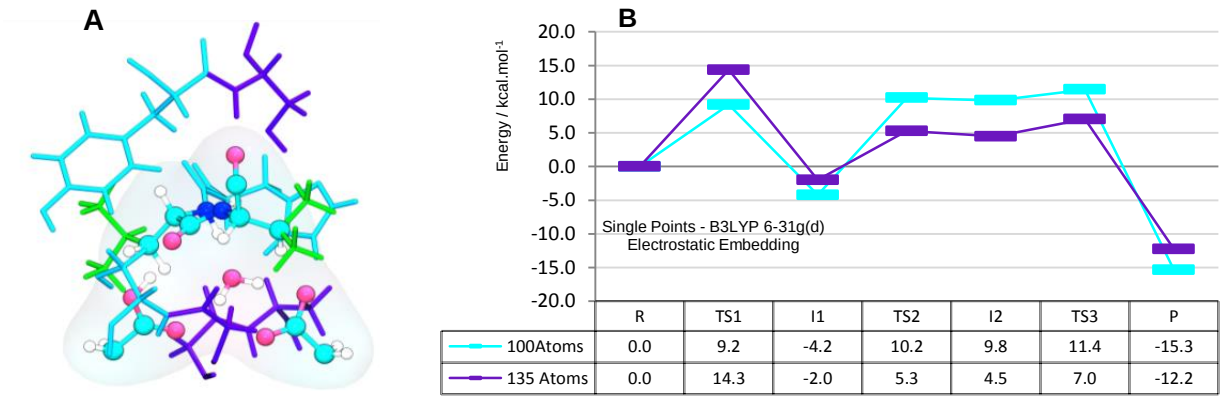


	R	TS1	I1	TS2	I2	TS3	P
55 Atoms	0.0	13.9	-3.8	8.9	8.2	11.5	-6.7
100Atoms	0.0	9.2	-4.2	10.2	9.8	11.4	-15.3

Figure 29. Single Point QM/MM calculations for a model with 100 atoms in high layer
A) Representation of the high layer residues. B) Energetic pathway along the reaction (blue representation).

In a third model Gly40, Ser84, Gly217 and Ala218 were added to the previous model. This new model has 135 atoms in the QM region. The calculated activation barrier for the

first TS was 13.1 kcal.mol^{-1} and this first step is exergonic by -2.0 kcal.mol^{-1}. The second TS had an activation barrier of 5.3 kcal.mol^{-1} and it is endergonic by 4.5 kcal.mol^{-1}. The third TS is associated with 7.0 kcal.mol^{-1} and the formation of the products is exergonic by -12.2 kcal.mol^{-1} (Figure 30).

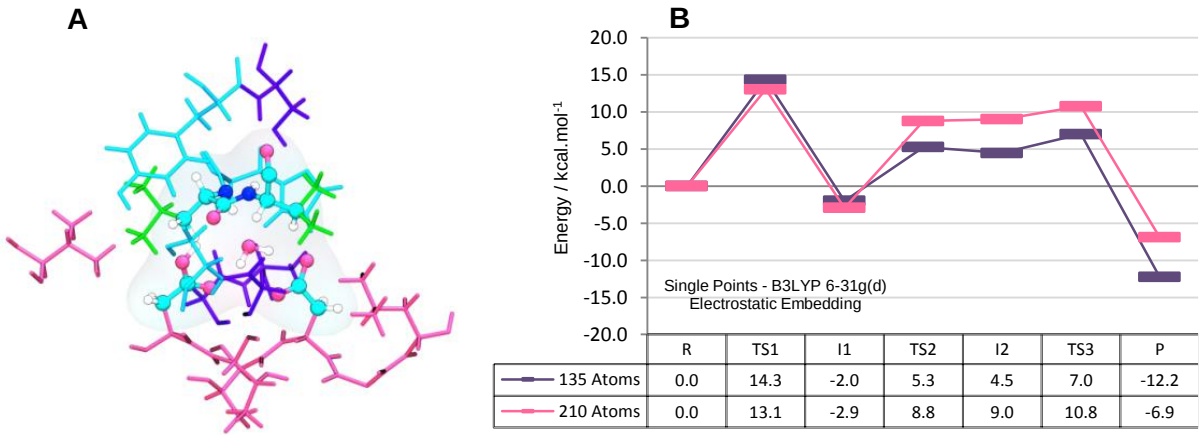


	R	TS1	I1	TS2	I2	TS3	P
100Atoms	0.0	9.2	-4.2	10.2	9.8	11.4	-15.3
135 Atoms	0.0	14.3	-2.0	5.3	4.5	7.0	-12.2

Figure 30. Single Point QM/MM calculations for a model with 135 atoms in high layer

A) Representation of the high layer residues. B) Energetic pathway along the reaction (violet representation).

We have tested also the influence of other atoms around the active site (Thr36, Val121, Leu213, Val214 and Thr216). This new system has 210 atoms in the high layer. The calculated activation barrier for the first TS was 13.1 kcal.mol^{-1} and this mechanistic step was exergonic by -2.9 kcal.mol^{-1}.



	R	TS1	I1	TS2	I2	TS3	P
135 Atoms	0.0	14.3	-2.0	5.3	4.5	7.0	-12.2
210 Atoms	0.0	13.1	-2.9	8.8	9.0	10.8	-6.9

Figure 31. Single Point QM/MM calculations for a model with 210 atoms in high layer

A) Representation of the high layer residues. B) Energetic pathway along the reaction (pink representation).

The second TS has an activation barrier of 8.8 kcal.mol^{-1} and it is very similar to the energy calculated for the second intermediate (9.0 kcal.mol^{-1}). This results may indicate that this second TS cannot correspond to a real TS or it was not correctly optimized. The third TS has a barrier of 10.8 kcal.mol^{-1} and the formation of the products is exergonic by -6.9 kcal.mol^{-1} (Figure 31).

Finally we expand our system to 322 atoms in QM layer. This new model includes the Leu75, part of Arg76, Gly80, Val122, Lys247, Thr289 and Ile330. The first TS has a barrier of 13.8 kcal.mol^{-1} and is little exergonic by -0.9 kcal.mol^{-1}. The second TS has a barrier of 14.2 kcal.mol^{-1} and it is endergonic by 13.7 kcal.mol^{-1}. The third TS has a barrier of 16.9 kcal.mol^{-1} and the products formation is little endergonic by 0.5 kcal.mol^{-1} (Figure 32).

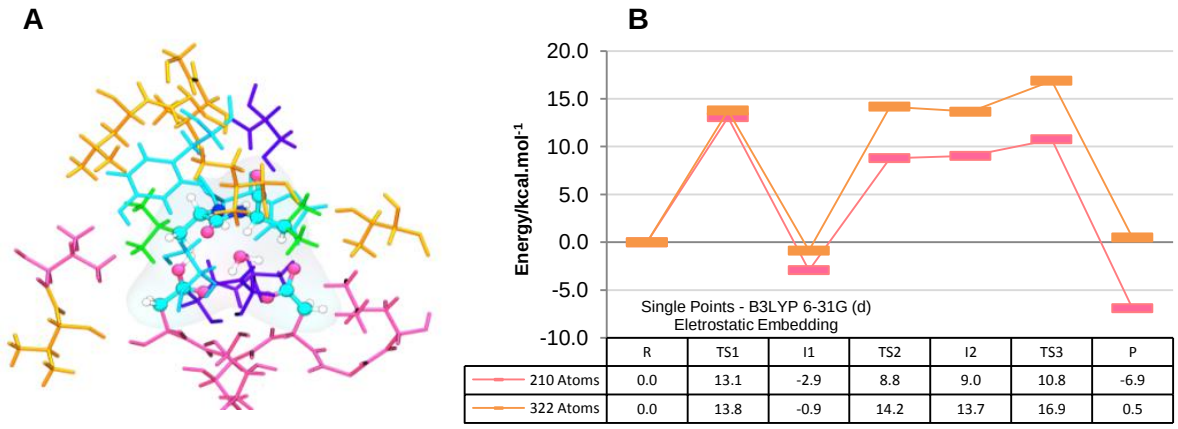


Figure 32. Single Point QM/MM calculations for a model with 322 atoms in high layer
A) Representation of the high layer residues. B) Energetic pathway along the reaction (orange representation).

The comparison between all these models shows some differences between the presented results. First of all, when the electrostatic embedding scheme is included in our calculation, the energies associated with each stationary point decrease in comparison with the energies obtained in optimizations with mechanical embedding scheme. These results were expected because the inclusion of the MM atoms polarization effect on the QM atoms, allows a better description of the system. Furthermore, when the QM region was expanded the relative energies decrease again. This means that the residues around the active site have a meaningful influence in reaction energy.

If we focus only in the first mechanistic step, which corresponds to the formation of the gem-diol intermediate, we can see that when the lateral chains of Val10 and Leu11 are included, the energetic barrier for the TS formation decrease about 3 kcal.mol^{-1}, which can

be indicative that this lateral chains are important for this first reaction step. When we expand the QM region the energy associated to this first step keeps approximately constant between the tested models. This is indicative that, probably, this influence is confined to the residues that contribute directly on the reaction.

However, the analysis of the second and third mechanistic steps shows some variances between the different models. These differences indicate that the added residues influence the energy associated with this second and third reactions. In the second step, the proton from Asp256 is transferred to the amide nitrogen of the scissile bond. Therefore, the Asp256 residue has a negative charge, and the amine nitrogen has a positive charge. These charges should interact strongly with the residues around the active site, and they can justify the differences discussed above.

4.6.2.3 Single point calculations with a large basis set

Afterwards the energies were recalculated at the B3LYP/6-311++G(2d,2p) level, also with the electrostatic embedding scheme. This large basis set was used to include the dispersive and polarization effects on the single point energy calculation.

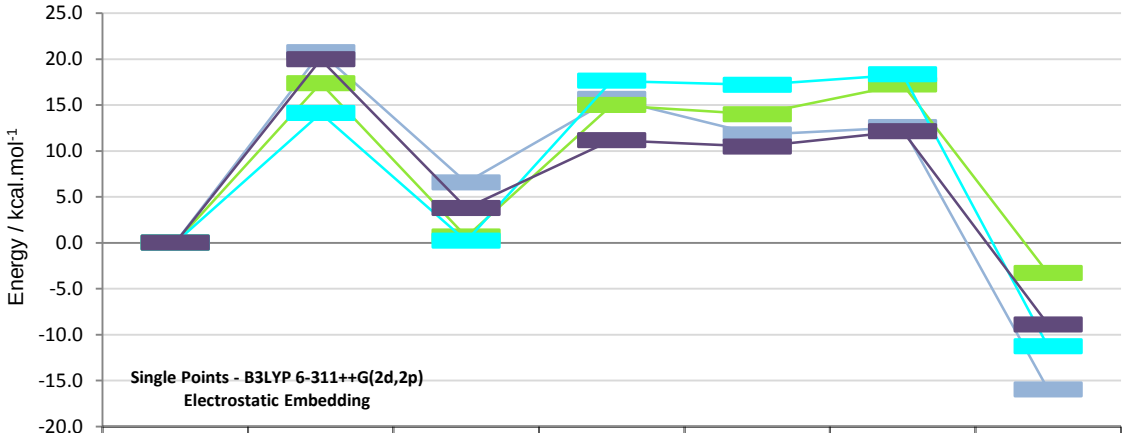
Using the initial QM region (33 atoms), the results indicated an energetic barrier of 20.7 kcal.mol^{-1} associated with the first TS and 6.6 kcal.mol^{-1} for the first mechanistic step. The second mechanistic step was also endergonic by 11.8 kcal.mol^{-1} with 15.7 kcal.mol^{-1} associated with the second TS formation. The barrier associated with the third mechanistic step was 12.6 kcal.mol^{-1} and the formation of the products was exergonic by 16.0 kcal.mol^{-1} (Figure 33 – grey representation).

In the model with 55 atoms in the QM region, the energetic barrier for the formation of the first TS was 17.4 kcal.mol^{-1} and the first TS is little endergonic by 0.7 kcal.mol^{-1}. The second and third transition states are associated with 15.0 kcal.mol^{-1} and 17.2 kcal.mol^{-1}, respectively and the reaction is exergonic by 3.3 kcal.mol^{-1} (Figure 33 – green representation).

In the model with 100 atoms in QM region, the first TS is associated with a barrier of 14.1 kcal.mol^{-1}. The second and third transition states are associated with 17.6 kcal.mol^{-1}, and 18.3 kcal.mol^{-1} respectively. The reaction is exergonic by -11.3 kcal.mol^{-1} (Figure 33 – blue representation).

In the model with 135 atoms, the first TS is associated with an energetic barrier of 20.0 kcal.mol^{-1}. The second and third transition states have a barrier of 11.2 kcal.mol^{-1} and 12.2 kcal.mol^{-1}. The reaction is exergonic by 8.9 kcal.mol^{-1} (Figure 33 – violet representation). The other two models (with 210 atoms and 322 atoms in high layer) were not tested because of the computational power that they need.

With these results it is possible to affirm that the first step of the reaction is probably the limiting one and the other two steps have very similar energies.



	R	TS1	I1	TS2	I2	TS3	P
33 Atoms	0.0	20.7	6.6	15.7	11.8	12.6	-16.0
55 Atoms	0.0	17.4	0.7	15.0	14.0	17.2	-3.3
100 Atoms	0.0	14.1	0.2	17.6	17.2	18.3	-11.3
135 Atoms	0.0	20.0	3.8	11.2	10.5	12.2	-8.9

Figure 33. Single Point QM/MM calculations with a large basis set (6-311++G(2d,2p))

The different colors represent the different models with 33, 55,100 and 135 atoms in high layer. The relative energies are represented in kcal.mol^{-1}and they were obtained with single point energy calculation at B3LYP/6-311++G(2d,2p)//Amber level.

This increase of the QM models was intended to include the effects of residues around the active center in the energy calculation for the stationary points. Even though much of the energy values calculated for reaction activated barrier stand at around 17-20 kcal.mol^{-1}, which is quite similar to the experimental energy associated with renin mechanism, there is some lack of agreement between the different studied models. Such observation may be due to the fact that the geometries of the stationary states have been taken directly from the PES scan and they are only approximate geometries for TS and INT points of the reaction. Probably they don't correspond to the very precise TS or INT and they require a re-optimization of the structure. Another explanation to these differences lies in the fact that the

added residues were previously optimized with MM and molecular embedding scheme, during the PES scan, and they were not re-optimized within the new level of theory.

The increasing of the QM model, only for energy calculations, requires considerable attention, not only in the choice of the added atoms, but also in the methods and in the interpretation of the results. However it is useful to understand the influence of the residues around the active site in a given catalytic reaction.

It is also important to note that the energetic problems observed in this study may also be due to the initial conformation of the enzyme. If a different initial conformation had been chosen as initial conformation, the calculated energy barriers could be different.

4.7 Human Renin vs. Mouse Renin

As mentioned previously, the mouse renin catalytic mechanism was previous described by our group [4]. The methodology used in these work was similar to those used in the present work. According to previous studies there are few differences between mouse renin and human renin. These two enzymes have about 68% of sequence identity [107] and they have similar folds and similar active site residues [108].

Mouse is the usual animal model to test new drugs and, therefore, it is very important to compare the mouse renin mechanism with human renin mechanism, in order to understand if the studies in mice are transferable to humans.

The comparison of mouse and human renin has revealed that these two enzymes have similar catalytic mechanism. The results of mouse renin work showed that the renin's mechanism of action proceeds through three mechanistic steps. First, the water molecule performs a nucleophilic attack to the carbonyl carbon of the scissile bond generating a gem-diol intermediate. In the second step the catalytic Asp residue (protonated in the previous step), protonates the amide nitrogen of the scissile bond, and only in a third step the cleavage of the peptide bond occurs, leading to the formation of two separate peptides. The results described in the present work have shown that human renin cleaves its substrate by the three same reaction steps. Furthermore the geometries of the stationary points (transition states, intermediates and products) for mouse and human renin were similar during all reaction.

These results show that both human and mouse enzymes appear to cleave the substrate by the same mechanism and with the same geometries along the reaction.

CHAPTER 5

CONCLUSIONS AND FUTURE PERSPECTIVES

5.1 Conclusions

Throughout this work we were able to describe the reaction path for angiotensinogen hydrolysis by human renin, with atomistic detail. The obtained results showed that this mechanism occurs by three steps, contrarily to other aspartic proteases. First, the enzyme activates the catalytic water molecule, which performs a nucleophile attack to the carbonyl carbon of the scissile angiotensinogen bond. This attack is accompanied by the deprotonation of the water molecule by Asp226 and, simultaneously by the protonation of the carbonyl oxygen by Asp38, generating a gem-diol intermediate. In the second step, Asp226 protonates the amide nitrogen of the substrate scissile bond. At this stage, this bond is weakened and elongated; however it only breaks in the third and last reaction step. In this last step, Asp38 deprotonates a hydroxyl group of the gem-diol and, simultaneously the other hydroxyl group is deprotonated by the other catalytic Asp residue. The human renin reaction leads, therefore, to the formation of two separate peptides, as well as two protonated catalytic aspartate residues.

It is also possible to conclude that the first mechanistic step of the hydrolysis reaction is rate limiting and it has an activation barrier of 20.0 kcal.mol^{-1} at B3LYP/6-311++G(2d,2p):Amber level of theory, with 135 atoms in high layer, which was the result obtained with the best theoretical level. This value is comparable with the experimental one

for the angiotensinogen terminal tetradecapeptide which is 16.5 kcal.mol^{-1}. The differences observed between the values, may be due to the fact that we use a dodecapeptide instead of a tetradecapeptide as substrate. Furthermore, single point energy calculations were used to re-calculate the energy associated with each stationary point, instead of a re-optimization of the structures with the new theoretical level, which may also have some influence on the obtained values. Additionally only an initial conformation was studied, instead of different initial conformations.

The comparison of the results obtained with electrostatic embedding and mechanical embedding schemes shows that the first provides a better description of the energetic profile of human renin reaction.

The obtained results also allow us to conclude that the structural water molecule, which is common in the most aspartic proteases, seems to be indifferent to the rearrangement of the active site during the renin catalytic mechanism.

With this work it was also possible to conclude that the catalysis of the mutated substrate (Leu10Phe) appears to be quite similar to the catalysis of the natural substrate. Accordingly, we can state that the association of preeclampsia to this mutation should not be explained by a change in the catalytic mechanism of the enzyme and it must be due to another reason.

Research on the renin catalytic mechanism is essential because this enzyme has an important role in the blood pressure control. The description of its mechanism of action and its transition states structures are the key for anti-hypertensive rational drug design. The transition state geometry can be taken as a template for the search of new transition-state analogues inhibitors. Such geometry is very difficult to obtain by experimental works, and because of it, the computational works are essential in this area.

This QM/MM study allows for a complete atomistic comprehension of renin catalytic mechanism and it may be useful to improve the efficiency of anti-hypertensive drug design.

5.2 Future Perspectives

With the results obtained in this work, it was possible to describe the catalytic mechanism of human renin. Last year, our group has published an article describing the catalytic mechanism of mouse renin. However, until now, there were no solid works

describing the catalytic mechanism of human renin. In order to clarify the detail of this mechanism, a paper with the results obtained in this work, is being prepared.

In attempt to improve the results obtained in this work, especially in terms of energy, the stationary points could be re-optimized without constraints. The structures obtained from this sort of optimization could be used to calculate vibrational frequencies, in order to compute zero-point vibration and entropic and thermal energy corrections to total energies.

In addition, the mechanism could be repeated using electrostatic embedding scheme and with different initial conformations, in order to compare with the obtained results. Furthermore an optimization of the models with large QM layers could improve the calculated energies, and probably it decreases the differences observed between the different models.

Although there are still issues for improvement, the information contained in this work is very important for the comprehension of renin catalytic mechanism and for the rational design of new renin inhibitors.

References

1. Ezzati, M., A.D. Lopez, A. Rodgers, S. Vander Hoorn, C.J.L. Murray, and C.R.A. Coll, *Selected major risk factors and global and regional burden of disease.* Lancet, 2002. **360**(9343): p. 1347-1360.

2. Bezencon, O., D. Bur, T. Weller, S. Richard-Bildstein, L. Remen, T. Sifferlen, O. Corminboeuf, C. Grisostomi, C. Boss, L. Prade, S. Delahaye, A. Treiber, P. Strickner, C. Binkert, P. Hess, B. Steiner, and W. Fischli, *Design and preparation of potent, nonpeptidic, bioavailable renin inhibitors.* J Med Chem, 2009. **52**(12): p. 3689-702.

3. Eder, J., U. Hommel, F. Cumin, B. Martoglio, and B. Gerhartz, *Aspartic proteases in drug discovery.* Curr Pharm Des, 2007. **13**(3): p. 271-85.

4. Brás, N.F., M.J. Ramos, and P.A. Fernandes, *The catalytic mechanism of mouse renin studied with QM/MM calculations.* Physical Chemistry Chemical Physics, 2012. **14**(36): p. 12605-12613.

5. Organization, W.H., *World health report 2001: mental health: new understanding, new hope*. 2001: World Health Organization.

6. Kearney, P.M., M. Whelton, K. Reynolds, P. Muntner, P.K. Whelton, and J. He, *Global burden of hypertension: analysis of worldwide data.* Lancet, 2005. **365**(9455): p. 217-223.

7. Whelton, P.K., J. He, L.J. Appel, J.A. Cutler, S. Havas, T.A. Kotchen, E.J. Roccella, R. Stout, C. Vallbona, and M.C. Winston, *Primary prevention of hypertension.* JAMA: the journal of the American Medical Association, 2002. **288**(15): p. 1882-1888.

8. Mancia, G., R. Fagard, K. Narkiewicz, J. Redon, A. Zanchetti, M. Böhm, T. Christiaens, R. Cifkova, G. De Backer, and A. Dominiczak, *2013 ESH/ESC Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC).* Blood pressure, 2013(0): p. 1-86.

9. Touyz, R.M., *New insights into mechanisms of hypertension.* Current Opinion in Nephrology and Hypertension, 2012. **21**(2): p. 119-121.

10. Muller, D.N., H. Kvakan, and F.C. Luft, *Immune-related effects in hypertension and target-organ damage.* Current opinion in nephrology and hypertension, 2011. **20**(2): p. 113-117.

11. Abboud, F.M., S.C. Harwani, and M.W. Chapleau, *Autonomic neural regulation of the immune system: implications for hypertension and cardiovascular disease.* Hypertension, 2012. **59**(4): p. 755.

12. Higuchi, S., H. Ohtsu, H. Suzuki, H. Shirai, G. Frank, and S. Eguchi, *Angiotensin II signal transduction through the AT1 receptor: novel insights into mechanisms and pathophysiology.* clinical Science, 2007. **112**: p. 417-428.

13. MacGregor, G., N. Markandu, J. Roulston, J. Jones, and J. Morton, *Maintenance of blood pressure by the renin–angiotensin system in normal man.* 1981.

14. Tigerstedt, R. and P. Bergman, *Niere und Kreislauf1.* Skandinavisches Archiv für Physiologie, 1898. **8**(1): p. 223-271.

15. Lopez, M.L.S.S. and R.A. Gomez, *Novel mechanisms for the control of renin synthesis and release.* Current hypertension reports, 2010. **12**(1): p. 26-32.

16. Verdecchia, P., F. Angeli, G. Mazzotta, G. Gentile, and G. Reboldi, *The renin angiotensin system in the development of cardiovascular disease: role of aliskiren in risk reduction.* Vascular health and risk management, 2008. **4**(5): p. 971.

17. Wu, C., H. Lu, L.A. Cassis, and A. Daugherty, *Molecular and pathophysiological features of angiotensinogen: a mini review.* North American journal of medicine & science, 2011. **4**(4): p. 183.

18. Katsurada, A., Y. Hagiwara, K. Miyashita, R. Satou, K. Miyata, N. Ohashi, L.G. Navar, and H. Kobori, *Novel sandwich ELISA for human angiotensinogen.* American Journal of Physiology-Renal Physiology, 2007. **293**(3): p. F956-F960.

19. Cumin, F., D. Le-Nguyen, B. Castro, J. Menard, and P. Corvol, *Comparative enzymatic studies of human renin acting on pure natural or synthetic substrates.* Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology, 1987. **913**(1): p. 10-19.

20. Campbell, D.J., *Angiotensin II generation in vivo: does it involve enzymes other than renin and angiotensin-converting enzyme?* Journal of Renin-Angiotensin-Aldosterone System, 2012. **13**(2): p. 314-316.

21. Campbell, D.J., T. Alexiou, H.D. Xiao, S. Fuchs, M.J. McKinley, P. Corvol, and K.E. Bernstein, *Effect of reduced angiotensin-converting enzyme gene expression and angiotensin-converting enzyme inhibition on angiotensin and bradykinin peptide levels in mice.* Hypertension, 2004. **43**(4): p. 854-859.

22. Crowley, S.D. and T.M. Coffman, *Recent advances involving the renin–angiotensin system.* Experimental cell research, 2012. **318**(9): p. 1049-1056.

23. Gainer, J.V., J.D. Morrow, A. Loveland, D.J. King, and N.J. Brown, *Effect of bradykinin-receptor blockade on the response to angiotensin-converting–enzyme inhibitor in normotensive and hypertensive subjects.* New England Journal of Medicine, 1998. **339**(18): p. 1285-1292.

24. Shen, X.Z., S. Billet, C. Lin, D. Okwan-Duodu, X. Chen, A.E. Lukacher, and K.E. Bernstein, *The carboxypeptidase ACE shapes the MHC class I peptide repertoire.* Nature immunology, 2011. **12**(11): p. 1078-1085.

25. Chappel, M. and C. Ferrario, *ACE and ACE2: their role to balance the expression of angiotensin II and angiotensin-(1–7).* Kidney international, 2006. **70**(1): p. 8-10.

26. Donoghue, M., F. Hsieh, E. Baronas, K. Godbout, M. Gosselin, N. Stagliano, M. Donovan, B. Woolf, K. Robison, and R. Jeyaseelan, *A novel angiotensin-converting enzyme–related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9.* Circulation research, 2000. **87**(5): p. e1-e9.

27. Timmermans, P.B., P. Benfield, A.T. Chiu, W.F. Herblin, P.C. Wong, and R.D. Smith, *Angiotensin II receptors and functional correlates.* American journal of hypertension, 1992. **5**(12 Pt 2): p. 221S-235S.

28. Timmermans, P. and R. Smith, *Angiotensin II receptor subtypes: selective antagonists and functional correlates.* European heart journal, 1994. **15**(suppl D): p. 79-87.

29. Dinh, D.T., A.G. Frauman, C.I. Johnston, and M.E. Fabiani, *Angiotensin receptors: distribution, signalling and function.* Clinical Science, 2001. **100**(5): p. 481-492.

30. McCarthy, C.A., R.E. Widdop, K.M. Denton, and E.S. Jones, *Update on the Angiotensin AT2 Receptor.* Current hypertension reports, 2013. **15**(1): p. 25-30.

31. Kaschina, E. and T. Unger, *Angiotensin AT1/AT2 receptors: regulation, signalling and function.* Blood pressure, 2003. **12**(2): p. 70-88.

32. Bader, M., *Tissue renin-angiotensin-aldosterone systems: targets for pharmacological therapy.* Annual review of pharmacology and toxicology, 2010. **50**: p. 439-465.

33. Schiffrin, E.L., *Circulatory therapeutics: use of antihypertensive agents and their effects on the vasculature.* Journal of cellular and molecular medicine, 2010. **14**(5): p. 1018-1029.

34. Bielecka-Dabrowa, A., W. S Aronow, J. Rysz, and M. Banach, *Current place of beta-blockers in the treatment of hypertension.* Current Vascular Pharmacology, 2010. **8**(6): p. 733-741.

35. Atlas, S.A., *The renin-angiotensin aldosterone system: pathophysiological role and pharmacologic inhibition.* Journal of managed care pharmacy, 2007. **13**(8): p. S9.

36. Bakris, G.L., J.B. Copley, N. Vicknair, R. Sadler, and S. Leurgans, *Calcium channel blockers versus other antihypertensive therapies on progression of NIDDM associated nephropathy.* Kidney international, 1996. **50**(5): p. 1641-1650.

37. Michel, J.-B., F.-X. Galen, C. Guettier, S. Sayah, N. Hinglais, J.-A. Fehrentz, D. le N'guyen, C. Carelli, B. Castro, and P. Corvol, *Immunological approach to blockade of the renin-substrate reaction.* Journal of hypertension. Supplement: official journal of the International Society of Hypertension, 1989. **7**(2): p. S63.

38. Pilz, B., E. Shagdarsuren, M. Wellner, A. Fiebeler, R. Dechend, P. Gratze, S. Meiners, D.L. Feldman, R.L. Webb, and I.M. Garrelds, *Aliskiren, a human renin inhibitor, ameliorates cardiac and renal damage in double-transgenic rats.* Hypertension, 2005. **46**(3): p. 569-576.

39. Wong, J., R.A. Patel, and P.R. Kowey, *The clinical use of angiotensin-converting enzyme inhibitors.* Progress in cardiovascular diseases, 2004. **47**(2): p. 116-130.

40. Hilgers, K.F. and J.F. Mann, *ACE inhibitors versus AT1 receptor antagonists in patients with chronic renal disease.* Journal of the American Society of Nephrology, 2002. **13**(4): p. 1100-1108.

41. Eder, J., U. Hommel, F. Cumin, B. Martoglio, and B. Gerhartz, *Aspartic proteases in drug discovery.* Current pharmaceutical design, 2007. **13**(3): p. 271-285.

42. Sielecki, A.R., K. Hayakawa, M. Fujinaga, M. Murphy, M. Fraser, A.K. Muir, C.T. Carilli, J.A. Lewicki, J.D. Baxter, and M. James, *Structure of recombinant human renin, a target for cardiovascular-active drugs, at 2.5 A resolution.* Science, 1989. **243**(4896): p. 1346-1351.

43. Nakagawa, T., J. Akaki, R. Satou, M. Takaya, H. Iwata, A. Katsurada, K. Nishiuchi, Y. Ohmura, F. Suzuki, and Y. Nakamura, *The His-Pro-Phe motif of angiotensinogen is a crucial determinant of the substrate specificity of renin.* Biological chemistry, 2007. **388**(2): p. 237-246.

44. Gradman, A.H. and R. Kad, *Renin inhibition in hypertension.* Journal of the American College of Cardiology, 2008. **51**(5): p. 519-528.

45. Brás, N.F., P.A. Fernandes, and M.J. Ramos, *Molecular dynamics studies on both bound and unbound renin protease.* Journal of Biomolecular Structure and Dynamics, 2013(ahead-of-print): p. 1-13.

46. Webb, R.L., N. Schiering, R. Sedrani, and J.r. Maibaum, *Direct renin inhibitors as a new therapy for hypertension.* 2010.

47. Cooper, J., *Aspartic proteinases in disease: a structural perspective.* Current drug targets, 2002. **3**(2): p. 155-173.

48. Sibanda, B.L., T. Blundell, P.M. Hobart, M. Fogliano, J.S. Bindra, B.W. Dominy, and J.M. Chirgwin, *Computer graphics modelling of human renin: Specificity, catalytic activity and intron-exon junctions.* FEBS letters, 1984. **174**(1): p. 102-111.

49. Hong, L. and J. Tang, *Flap position of free memapsin 2 (β-secretase), a model for flap opening in aspartic protease catalysis.* Biochemistry, 2004. **43**(16): p. 4689-4695.

50. Barman, A. and R. Prabhakar, *Elucidating the catalytic mechanism of β-secretase (BACE1): A quantum mechanics/molecular mechanics (QM/MM) approach.* Journal of Molecular Graphics and Modelling, 2013. **40**: p. 1-9.

51. Cascella, M., C. Micheletti, U. Rothlisberger, and P. Carloni, *Evolutionarily conserved functional mechanics across pepsin-like and retroviral aspartic proteases.* Journal of the American Chemical Society, 2005. **127**(11): p. 3734-3742.

52. Garrec, J., P. Sautet, and P. Fleurat-Lessard, *Understanding the HIV-1 protease reactivity with DFT: what do we gain from recent functionals?* The Journal of Physical Chemistry B, 2011. **115**(26): p. 8545-8558.

53. Singh, R., A. Barman, and R. Prabhakar, *Computational insights into aspartyl protease activity of presenilin 1 (PS1) generating Alzheimer amyloid β-peptides (Aβ40 and Aβ42).* The Journal of Physical Chemistry B, 2009. **113**(10): p. 2990-2999.

54. Andreeva, N.S. and L.D. Rumsh, *Analysis of crystal structures of aspartic proteinases: On the role of amino acid residues adjacent to the catalytic site of pepsin-like enzymes.* Protein Science, 2001. **10**(12): p. 2439-2450.

55. Staessen, J.A., Y. Li, and T. Richart, *Oral renin inhibitors.* The Lancet, 2006. **368**(9545): p. 1449-1456.

56. Wood, J.M., N. Gulati, P. Forgiarini, W. Fuhrer, and K.G. Hofbauer, *Effects of a specific and long-acting renin inhibitor in the marmoset.* Hypertension, 1985. **7**(5): p. 797-803.

57. Wood, J.M., F. Cumin, and J. Maibaum, *Pharmacology of renin inhibitors and their application to the treatment of hypertension.* Pharmacology & therapeutics, 1994. **61**(3): p. 325-344.

58. Wood, J.M., J. Maibaum, J. Rahuel, M.G. Grütter, N.-C. Cohen, V. Rasetti, H. Rüger, R. Göschke, S. Stutz, and W. Fuhrer, *Structure-based design of aliskiren, a novel orally effective renin inhibitor.* Biochemical and biophysical research communications, 2003. **308**(4): p. 698-705.

59. Jensen, C., P. Herold, and H.R. Brunner, *Aliskiren: the first renin inhibitor for clinical treatment.* Nature reviews Drug discovery, 2008. **7**(5): p. 399-410.

60. Ramos, M.J. and P.A. Fernandes, *Computational enzymatic catalysis.* Accounts of chemical research, 2008. **41**(6): p. 689-698.

61. Lonsdale, R., J.N. Harvey, and A.J. Mulholland, *A practical guide to modelling enzyme-catalysed reactions.* Chemical Society Reviews, 2012. **41**(8): p. 3025-3038.

62. Sousa, S.F., P.A. Fernandes, and M.J. Ramos, *Computational enzymatic catalysis–clarifying enzymatic mechanisms with the help of computers.* Physical Chemistry Chemical Physics, 2012. **14**(36): p. 12431-12441.

63. Ramos, M.J. and P.A. Fernandes, *Atomic-level rational drug design.* Current Computer-Aided Drug Design, 2006. **2**(1): p. 57-81.

64. Gluza, K. and P. Kafarski, *Transition State Analogues of Enzymatic Reaction as Potential Drugs.* 2013.

65. Schrödinger, E., *An undulatory theory of the mechanics of atoms and molecules.* Physical Review, 1926. **28**(6): p. 1049.

66. Leach, A.R., *Molecular modelling: principles and applications*. 2001: Pearson Education.

67. Cramer, C.J., *Essentials of computational chemistry : theories and models*. 2nd ed. 2004, Chichester, West Sussex, England ; Hoboken, NJ: Wiley. xx, 596 p.

68. Jensen, F., *Introduction to computational chemistry*. 2007: Wiley. com.

69. Shaw, D.E., P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, and W. Wriggers, *Atomic-level characterization of the structural dynamics of proteins.* Science, 2010. **330**(6002): p. 341-6.

70. Young, D.C., *Computational chemistry : a practical guide for applying techniques to real world problems*. 2001, New York: Wiley. xxiv, 381 p.

71. Weiner, P.K. and P.A. Kollman, *AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions.* Journal of Computational Chemistry, 1981. **2**(3): p. 287-303.

72. Brooks, B.R., R.E. Bruccoleri, B.D. Olafson, S. Swaminathan, and M. Karplus, *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations.* Journal of computational chemistry, 1983. **4**(2): p. 187-217.

73. Jorgensen, W.L., D.S. Maxwell, and J. Tirado-Rives, *Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids.* Journal of the American Chemical Society, 1996. **118**(45): p. 11225-11236.

74. Jones, J., *On the determination of molecular fields. II. From the equation of state of a gas.* Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 1924. **106**(738): p. 463-477.

75. Alder, B.J. and T. Wainwright, *Studies in molecular dynamics. I. General method.* The Journal of Chemical Physics, 1959. **31**: p. 459.

76. McCamm0n, J.A., *Dynamics of folded proteins.* Nature, 1977. **267**: p. 16.

77. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules.* Nat Struct Biol, 2002. **9**(9): p. 646-52.

78. Born, M. and R. Oppenheimer, *Zur quantentheorie der molekeln.* Annalen der Physik, 1927. **389**(20): p. 457-484.

79. Thomas, L.H. *The calculation of atomic fields*. in *Mathematical Proceedings of the Cambridge Philosophical Society*. 1927. Cambridge Univ Press.

80. Fermi, E., *Un metodo statistico per la determinazione di alcune priorieta dell'atome.* Rend. Accad. Naz. Lincei, 1927. **6**(602-607): p. 32.

81. Sousa, S.F., P.A. Fernandes, and M.J. Ramos, *General performance of density functionals.* J Phys Chem A, 2007. **111**(42): p. 10439-52.

82. Hohenberg, P. and W. Kohn, *Inhomogeneous electron gas.* Physical review, 1964. **136**(3B): p. B864.

83. Kohn, W. and L.J. Sham, *Self-consistent equations including exchange and correlation effects.* Physical Review, 1965. **140**(4A): p. A1133.

84. Vreven, T., K.S. Byun, I. Komaromi, S. Dapprich, J.A. Montgomery, K. Morokuma, and M.J. Frisch, *Combining quantum mechanics methods with molecular mechanics methods in ONIOM.* Journal of Chemical Theory and Computation, 2006. **2**(3): p. 815-826.

85. Dapprich, S., I. Komáromi, K.S. Byun, K. Morokuma, and M.J. Frisch, *A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives.* Journal of Molecular Structure: THEOCHEM, 1999. **461**: p. 1-21.

86. Svensson, M., S. Humbel, R.D. Froese, T. Matsubara, S. Sieber, and K. Morokuma, *ONIOM: A multilayered integrated MO+ MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt (P (t-Bu) 3) 2+ H2 oxidative addition.* The Journal of Physical Chemistry, 1996. **100**(50): p. 19357-19363.

87. Vreven, T., K. Morokuma, O. Farkas, H.B. Schlegel, and M.J. Frisch, *Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints.* J Comput Chem, 2003. **24**(6): p. 760-9.

88. Groenhof, G., *Introduction to QM/MM simulations*, in *Biomolecular Simulations*. 2013, Springer. p. 43-66.

89. 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 Research.* 2000. **28**: p. 235-242.

90. Zhou, A., R.W. Carrell, M.P. Murphy, Z. Wei, Y. Yan, P.L. Stanley, P.E. Stein, F.B. Pipkin, and R.J. Read, *A redox switch in angiotensinogen modulates angiotensin release.* Nature, 2010. **468**(7320): p. 108-111.

91. Inoue, I., A. Rohrwasser, C. Helin, X. Jeunemaitre, P. Crain, J. Bohlender, R.P. Lifton, P. Corvol, K. Ward, and J.M. Lalouel, *A mutation of angiotensinogen in a patient with preeclampsia leads to altered kinetics of the renin-angiotensin system.* J Biol Chem, 1995. **270**(19): p. 11430-6.

92. Case D. A., T.A.D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo,, P.D.A. Merz K.M., M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I., K.F.W. Kolossváry, F. Paesani, J. Vanicek, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke,, C.T. L. Yang, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin,, and a.P.A. Kollman, *AMBER 9.* University of California, San Francisco, California, 2006.

93. Piana, S., D. Bucher, P. Carloni, and U. Rothlisberger, *Reaction mechanism of HIV-1 protease by hybrid Car-Parrinello/classical MD simulations.* The Journal of Physical Chemistry B, 2004. **108**(30): p. 11139-11149.

94. Frisch, M.J.T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, *Gaussian 09, Revision D.01.* 2009.

95. Case, D.A., T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R.J. Woods, *The Amber biomolecular simulation programs.* Journal of computational chemistry, 2005. **26**(16): p. 1668-1688.

96. Hammonds, K.D. and J.-P. Ryckaert, *On the convergence of the SHAKE algorithm.* Computer Physics Communications, 1991. **62**(2): p. 336-351.

97. Izaguirre, J.A., D.P. Catarello, J.M. Wozniak, and R.D. Skeel, *Langevin stabilization of molecular dynamics.* The Journal of chemical physics, 2001. **114**: p. 2090.

98. Becke, A.D., *Density-functional thermochemistry. III. The role of exact exchange.* The Journal of Chemical Physics, 1993. **98**: p. 5648.

99. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Physical Review B, 1988. **37**(2): p. 785.

100. Singh, U.C. and P.A. Kollman, *An approach to computing electrostatic charges for molecules.* Journal of Computational Chemistry, 1984. **5**(2): p. 129-145.

101. Besler, B.H., K.M. Merz, and P.A. Kollman, *Atomic charges derived from semiempirical methods.* Journal of Computational Chemistry, 1990. **11**(4): p. 431-439.

102. Bayly, C.I., P. Cieplak, W. Cornell, and P.A. Kollman, *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model.* The Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280.

103. Dunning Jr, T.H., *Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen.* The Journal of Chemical Physics, 1989. **90**: p. 1007.

104. Barman, A., S. Schürer, and R. Prabhakar, *Computational modeling of substrate specificity and catalysis of the β-secretase (BACE1) enzyme.* Biochemistry, 2011. **50**(20): p. 4337-4349.

105. Prasad, B. and K. Suguna, *Role of water molecules in the structure and function of aspartic proteinases.* Acta Crystallographica Section D: Biological Crystallography, 2002. **58**(2): p. 250-259.

106. Pilote, L., G. McKercher, D. Thibeault, and D. Lamarre, *Enzymatic characterization of purified recombinant human renin.* Biochemistry and cell biology, 1995. **73**(3-4): p. 163-170.

107. Imai, T., H. Miyazaki, S. Hirose, H. Hori, T. Hayashi, R. Kageyama, H. Ohkubo, S. Nakanishi, and K. Murakami, *Cloning and sequence analysis of cDNA for human renin precursor.* Proceedings of the National Academy of Sciences, 1983. **80**(24): p. 7405-7409.

108. Akahane, K., H. Umeyama, S. Nakagawa, I. Moriguchi, S. Hirose, K. Iizuka, and K. Murakami, *Three-dimensional structure of human renin.* Hypertension, 1985. **7**(1): p. 3-12.