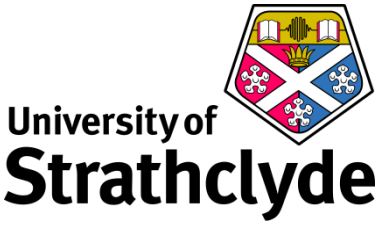


Molecular Design, GlaxoSmithKline

Department of Pure and Applied Chemistry

COMPUTATIONAL MODELLING OF ENZYME ACTIVITY TO SPEED UP BIOCATALYST REDESIGN

by

MARIE-PIERRE DRÉANIC

A thesis submitted to the Department of Pure and Applied Chemistry, University of Strathclyde, in part fulfillment of the regulations for the degree of Doctor of Philosophy in Chemistry.

April 2018

DECLARATION OF OWNERSHIP

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Date:

Acknowledgements

I would like to express my profound gratitude to my supervisors Colin Edge and Tell Tuttle. Colin for his everyday support and guidance, for his great availability to share his deep knowledge in computational chemistry with me, and for helping me improve my English writing skills; he has passed a lot of time in this respect.

Tell, even from distance, has provided very useful academic advice and guidelines. He helped me by proper guidance, effective comments and with good support. He also paved the way for me to conduct my PhD program.

Kris Brown, computational chemist at GlaxoSmithKline, for providing me case studies to work on.

James Woolven, computational chemist at GlaxoSmithKline, for sharing his expertise in Schrödinger software. He dedicated lot of time helping me solve various problems.

David Rinaldo, applications scientist at Schrodinger, for his precious guidelines regarding the use of Qsite and availability to solve any technical issue.

Peter Pogany, computational chemist at GlaxoSmithKline, for sharing his expertise in quantum methods and for providing me useful guidance.

Pamela Thomas, computational chemist at GlaxoSmithKline, for the useful discussions we had in structural bioinformatics.

I would also like to thank all past and present members of the GSK computational chemistry team for their advices, sympathy and good humour; they make my PhD a very enjoyable experience.

I would also like to thank my industrial PhD Programme Director, Professor William Kerr and Fellow Programme Director, Dr Harry Kelly, for making this Industrial PhD possible.

Schrodinger for the student licenses they provided for my PhD work.

Finally, my family for their continuous and unparalleled support and encouragement.

Abbreviations

6-APA	6-Aminopenicillanic Acid
AR	Aldose Reductase
CALB	*Candida Antarctica* Lipase B
CASP	Critical Assessment of methods of protein Structure Prediction
CPU	Central Processing Unit
DE	Directed Evolution
DFT	Density Functional Theory
DNA	Deoxyribonucleic Acid
E.U.	European Union
E_a	Activation Energy
EC	Enzyme Commission
ee	Enantiomeric Excess
EOC	Ethyl Oxazole-5-Carboxylate
EP	Enzyme-Product Complex
epPCR	Error-Prone Polymerase Chain Reaction
E_R	Relative Energy
ES	Enzyme-Substrate Complex

EVB	Empirical Valence Bond
FDA	Food and Drug Administration
FS	Full System
GGA	Generalized Gradient Approximation
GLD	D-Glyceraldehyde
GPU	Graphics Processing Unit
GSK	GlaxoSmithKline
ΔG	Gibbs Free Energy
HF	Hartree-Fock
HTS	High-Throughput-Screening
IL	Ionic Liquid
IPP	1-Isopropylpiperazine
IREDs	Imine Reductases
IS	Inner Region
ISM	Iterative Saturation Mutagenesis
KIE	Kinetic Isotope Effect
KREDs	Ketoreductases
KS	Kohn-Sham
LA	Link Atom

LCAO	Linear Combination of Atomic Orbitals
LDA	Local-Density Approximations
MAOs	Monoamine Oxidases
MBZ	Methyl Benzoate
MD	Molecular Dynamics
MM	Molecular Mechanics
MOE	Molecular Operating Environment
MW	Molecular Weight
NADP	Nicotinamide-Adenine-Dinucleotide Phosphate
NME	New Molecular Entities
NTD	New Therapeutic Drug
OS	Outer Region
PAL	*Pseudomonas Aeruginosa* Lipase
PBL	*Pseudomonas Batumici* Lipase
PDB	Protein Data Bank
PES	Potential Energy Surface
PLP	Pyridoxal-5'-Phosphate
PMA	Phenylmethanamine
QM	Quantum Mechanics

QM/MM	Quantum Mechanics/Molecular Mechanics
SSE	Secondary Structure Element
SSM	Site-Saturation Mutagenesis
TAms	Transaminases
T_m	Melting Temperature
TS	Transition State
TST	Transition State Theory
U.S.	United States of America
XC	Exchange-Correlation

Abstract

Biocatalysis is increasingly used for the synthesis of pharmaceuticals intermediates. However, to expand the applicability of these methods current timelines for biocatalyst optimization need to be reduced. Quantum Mechanics/Molecular Mechanics (QM/MM) methods allow, in principle, the accurate evaluation of enzymatic activities and thus offer an interesting option for the *in silico* pre-screening of variants. However, standard QM/MM methods are a computationally expensive class of methods and thus for practical large scale applications, approximations need to be made. In this work, a QM/MM-based protocol for hotspot identification has been developed and tested.

The establishment and validation of an internal protocol for accurate QM/MM calculations was achieved through the mechanistic study of aldose reductase. This study highlights the importance of parameters, such as size of the QM region or the choice of the QM method, on differentiating between competitive mechanisms and consequently on accurately determining the role of the environment on the energy profile of the reaction.

Having a validated QM/MM methodology, an enzymatic amide bond formation was used as a case study to elaborate and test a protocol for hotspot identification. In the resulting protocol three major approximations were introduced to speed up calculations: starting from a single snapshot, neglecting possible reorganizations consecutive to the mutation and focusing mainly on electrostatic effects. Two different types of charge modification protocols were investigated: charge deletion and charge introduction. From this study one specific hotspot was identified.

In a further study, a homology model strategy was conducted to cope with the absence of experimentally determined structure, a frequent issue in enzyme design. Our previously established protocol was re-tested starting from the homology model and hotspots were identified. Finally, an evaluation of solvent free calculations, as an option to further accelerate the calculations, was also carried out. Encouraging results were obtained in the solvent free studies as similar hotspots were obtained relative to the water or toluene

solvated models. Nonetheless, significant variations do exist between different solvents and further studies are necessary to validate the use of this approximation in a wider context.

Table of Contents

1. Introduction

1.1 Motivation of the Thesis

1.1.1 The long and costly drug discovery and development process

The pharmaceutical industry brings life-saving and life-enhancing medicines to patients. From a patient's point of view there seems to be an unending stream of innovative products. From 1950 to 2008, the US Food and Drug Administration (FDA) approved 1,222 new drugs (new molecular entities (NMEs) or new biologics).[1] There is, however, a high cost that must be borne by the industry as the drug discovery and development process is long and expensive. In fact, for every approved new therapeutic drug (NTD) there is, on average, a total R&D cost of over $2.5 billion,[2] over an average of 14 year timescale for the research and development process associated with each NTD.[3] As such, new approaches that offer the potential to save time and money are actively sought within the industry. For example, technologies to reduce the timelines and increase success rates include the implementation of combinatorial chemistry, DNA sequencing, high-throughput-screening (HTS) or computational modelling.[4] This work focuses on computational technology to speed up the optimization of enzyme catalysts.

The pharmaceutical industry aims to bringing key medicines to the patient with minimum impact on the environment and enzymes are playing an increasing role in achieving a sustainable process.[5-6] However enzymes usually need to be engineered to suit the target process. Current timelines for engineering enzymes are considered not fast enough for a

perfect integration before product launch.[7] Computational tools could be an important driver for increased speed.

1.1.2 Computational modelling to improve efficiency

Computational modelling can be defined as the use of computers to simulate and study the behaviour of complex systems using mathematics, physics and computer science.[8] Experimentation is still required to help a researcher find solutions to problems, but computer modelling can reduce the time and cost by prioritising work.

Protein-ligand interactions (and more specifically enzyme-ligand interactions) are omnipresent along the length of the drug discovery and development pipeline (**Figure 1.1**) and thus accurate modelling of these interactions offers the possibility of accelerating several components of the process.

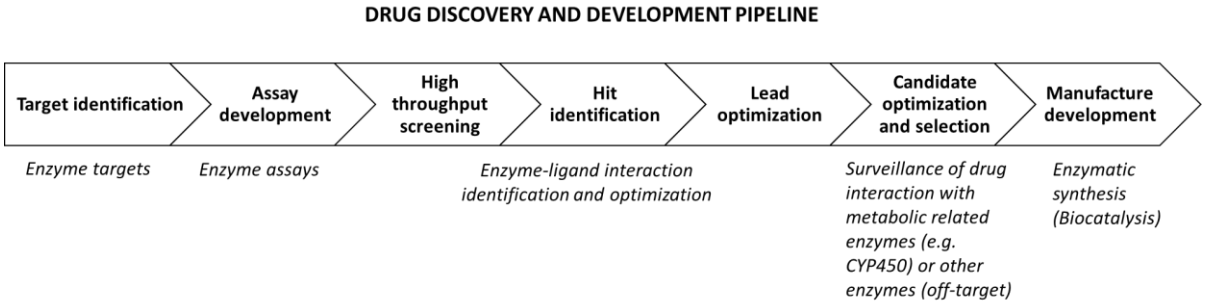


Figure 1.1. Example of enzyme-ligand systems along the drug discovery and development pipeline.

Computational chemistry is a branch of computational modelling used for the calculation of molecular properties or the simulation of molecular behavior.[9] Various computational

chemistry methods can be used to study protein-ligand interactions, and depending on their underlying methodology offer differing levels of predictive potential. The choice of the method is always a quest for the optimal accuracy/speed trade-off so that the time-scale remains competitive or faster than existing experimental approaches. In the pharmaceutical industry, the primary effort of a computational chemistry group is to support medicinal chemists in small-molecule drug discovery projects. Namely, to predict the affinity of a ligand for its protein target as this is directly related to the drug activity.

Recently, various major pharma companies throughout the world (Janssen, Boehringer Ingelheim, Roche, Bayer and Genentech) have described their use of computational chemistry during the drug discovery and development process and docking was systematically reported as the method of choice for target-ligand interaction estimation.[10-14] The importance of free energy perturbation (FEP),[15] a molecular dynamics (MD) based method, has also been highlighted. The gain in popularity of molecular dynamics is linked to the influence of Moore's Law.[16] According to this law the transistor count of CPUs approximately doubles every 2 years. The recent progress in software and hardware (*e.g.*, GPU clusters) now enable high speed MD simulations. In contrast, the use of quantum mechanics (QM) methods in pharma companies is more anecdotical (mainly for torsional analysis or pK_a predictions) and Quantum mechanics / Molecular mechanics (QM/MM) methods in particular are not reported in any of these papers.[10-14] The drawbacks of QM-based methods is neatly summarised in a recent article entitled "Application of Molecular Modelling to Speed-up the Lead Discovery Process" where the author writes that QM based methods are "too computationally expensive and applicable only to small molecular

systems".[17] Nevertheless, this very computationally intensive class of applications based on electronic structure calculations could also gain popularity by a correct exploitation of GPU acceleration.[18]

QM/MM techniques offer multiple advantage over classic molecular mechanical (MM) approaches. Firstly, they allow a description of the electronic structure for the active site of proteins. For example, a better representation can be made for polarization effects, dipole moment or cation–π and π–π interactions. QM/MM formalisms thus provide an opportunity to improve the accuracy of a diverse set of calculations such as docking or electrostatic calculations. Also, QM/MM approaches allow bond formation and/or breakage which is essential for the study of enzyme reactivity.

However, it is not always straightforward for a computational chemistry group to integrate new (academic) methodologies as this process is usually time consuming. In general, this will be done in collaboration with academic partners and/or via a student project (summer, industrial placement, PhD Case, Post-doc …).[10-11] This is why the computational chemistry department of GSK wanted to dedicate an "Industrial Ph.D." project to the evaluation of QM/MM methodologies through the GSK/Strathclyde scheme.[19] As such, the work described in this thesis is directly aligned to GSK business and discovery projects.

1.2 Layout of the Thesis

The work presented in this thesis describes the development of QM/MM protocols to rationalize experimental strategies in biocatalysis design. **Chapter 2** introduces the field of biocatalysis for the pharmaceutical industry. It describes the motivation for using biocatalysis in the pharmaceutical industry, the approaches that are used to tailor specific biocatalysts for a process and new concepts and methods that could help improve this process.

In **Chapter 3** the theoretical background needed to understand the work in this thesis is introduced. A particular focus is given to the theory underlying quantum mechanics/molecular mechanics methods with an emphasis on the methodology used in this work.

Chapter 4 describes the QM/MM mechanistic study of aldose reductase (AR). This key enzyme is important as it has been linked to some diabetes mellitus complications. The mechanism of the enzyme was investigated by using three different protonation states of the active site residue His110. The study of this "classic" system offers a validation of the QM/MM methodology used and also presents a modified version of the previously contentious mechanism.[20]

Chapter 5 reports the QM/MM mechanistic study of amide bond formation catalysed by thermophilic carboxylesterase and the hotspot identification for a target reaction. 'Amide bond formation avoiding poor atom economy reagent' is a top priority on-going challenge in the area of sustainable chemistry. Within GlaxoSmithKline an enzymatic option in the

synthetic route of a recently disclosed PI3Kδ inhibitor has been investigated. A carboxylesterase with an available crystal structure was selected for the project. Starting from this structure, QM/MM calculations were carried out by following the validated methodology from Chapter 4 and a QM/MM protocol for hotspot identification was then established and tested. Two charge modification protocols were investigated and hotspots identified from them.

Chapter 6 investigates the amide bond formation catalysed by a lipase and evaluates the effect of the solvent choice on the hotspot identification output. The target reaction in this study is part of an enzymatic panel for amide bond formation currently under development. Unlike the carboxylesterase from Chapter 5, the selected enzyme for this project was a lipase without any experimental structure. This chapter thus first describes the homology modelling work. The reaction mechanism is then investigated. Finally, the effect of the solvent environment on the reaction mechanism and on charge modification protocols results is investigated.

2. Enzyme Biocatalysis

2.1 Enzymes

Catalysis is the enhancement of the rate of a reaction due to the presence of an additional molecule that is not consumed during the reaction, the catalyst.[21] In biology, most chemical transformations are catalysed by enzymes.[22] Enzymes are essential to life by reducing the time scale of a variety of reactions from millions of years to fraction of seconds and thus allowing organisms to have a functional metabolism.[1] The Enzyme Commission (EC) of the International Union of Biochemistry classifies enzymes into six classes based on the type of chemical reaction they catalyse.[23-24]

Table 2.1. Description of the main Enzyme classes.

Enzyme classes	EC 1	EC 2	EC 3	EC 4	EC 5	EC 6
Name	Oxidoreductase	Transferase	Hydrolase	Lyase	Isomerase	Ligase

2.1.1 Structure

Enzymes are macromolecular catalysts and more precisely proteins. As with any other protein, enzymes are composed of chain(s) of amino acids. There are twenty gene-encoded amino acids abbreviated by a one or three letter code (**Figure 2.1**). Amino acids are chemically diverse building blocks whose mutual interactions determine both the three dimensional configuration and function of a protein.[25] The specific order of these amino

acids within a chain is called the protein sequence and defines the primary structure of the enzyme.

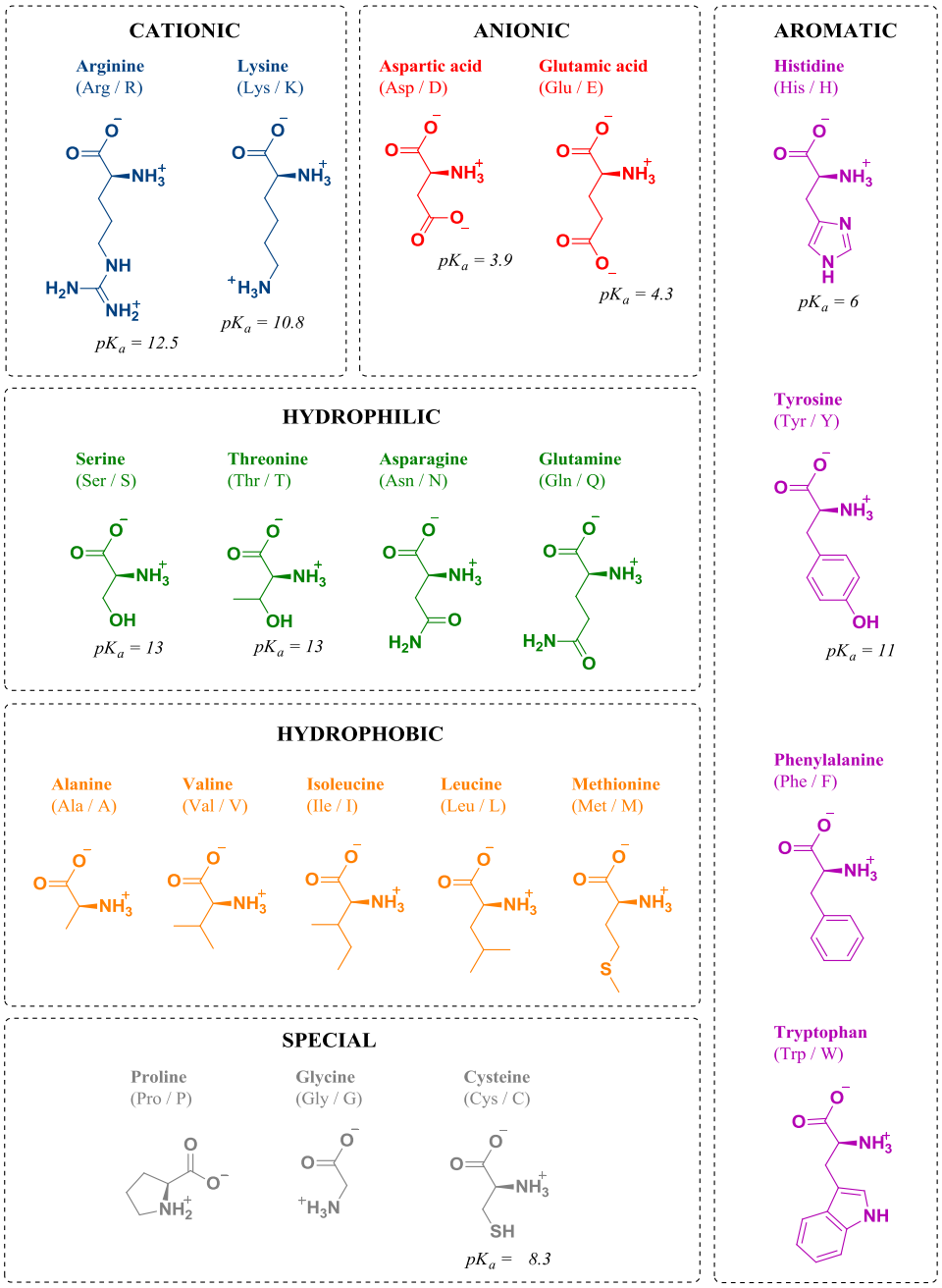


Figure 2.1. Amino acids classification. Protonation state shown at pH 7. pK_a values of polar and charged amino acids side chains shown in black are taken from [25].

The secondary structure is determined by the pattern of hydrogen bonding and defines the three-dimensional form of *local* segments of proteins. The most common secondary structure elements (SSE) are α-helices[26] and β-sheets[27] (**Figure 2.2**), they represent approximately 30% and 20%, respectively, of the SSE composition of proteins in the Protein Data Bank (PDB)[28].[29] Turns[30] (mostly β-turns[31]) are the third most common SSE which account for around 10% of the PDB secondary structure composition. There are various other SSE that are less frequent, for examples 3_{10}-helices[31] is the fourth most frequent SSE and accounts for 3%. The tertiary structure refers to the three-dimensional structure of the enzyme, it arises from the spatial arrangement of the SSE and is also known as the protein fold. This fold can be stabilised by both non-covalent (hydrogen bonds, charge-charge and hydrophobic interactions) and covalent interactions (disulphide bridges). Depending on the relative content of α-helices and β-sheets enzyme folds can be classified into three major classes — mainly α (*e.g.*, myoglobin), mainly β (*e.g.*, fatty acid binding protein) and α–β (*e.g.*, triosephosphate isomerase).[32] α–β can be further divided depending on whether α–helices and β-sheets are alternated (α/β) or largely-separated (α+β). Finally, assembly of multiple folded proteins can be formed and is referred to as quaternary structure.

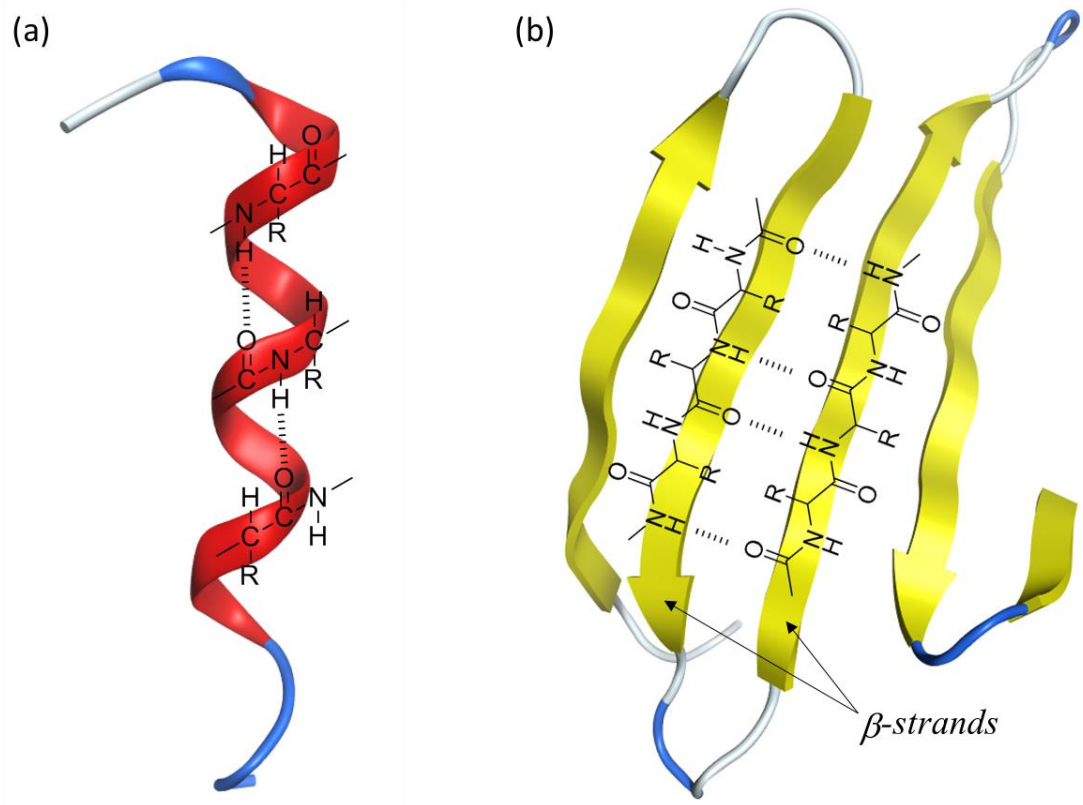


Figure 2.2. Examples of secondary structures (a) α-helix and (b) antiparallel β-sheet (made of β-strands).

Because enzyme three-dimensional structure is mainly held together by non-covalent interactions (*e.g.*, hydrogen bonds), higher temperatures, organics solvents or strong acids and bases can break those bonds and lead to the loss of the native structure, a process called denaturation.

In enzymes, catalysis takes place in a defined region of the tertiary structure, called the active site. In this region, the enzyme supplies functional groups, via the amino acids residues' side chains, necessary for the reaction.[25] The combination of these residues are

called ‘catalytic units’ and are usually form of two (dyad) or three (triad) residues.[25] One well know example is the Ser-His-Asp catalytic triad of α/β hydrolase fold enzymes.[33] The special arrangement of these residues, determined by the enzyme fold, is critical to the reaction and thus governs the enzyme function.

2.1.2 Function

The active site within an enzyme dictates the type of the reaction that the enzyme can catalyse. However, the activity of the enzyme is not uniquely dependent on this and is best described by the Michaelis-Menten model (**Eq. 2.1**).[34] From this equation it is evident that the activity of the enzyme depends on two main events: first the reversible formation of the enzyme-substrate complex (ES) and second a catalytic step where the ES complex is converted into an enzyme-product complex (EP) which is followed by the product (P) release. The conversion of the substrate into a product occurs with a catalytic rate constant k_{cat}, which measures the effectiveness of a reaction for enzymes working under substrate saturation. However, because enzyme catalysis can be limited by how well the substrate binds to the enzyme, other parameters are needed. K_M (Michaelis constant) is the dissociation constant of the ES complex defined as $(k_{-1}+k_{cat})/k_1$. The ratio k_{cat}/K_M is referred to as the specificity constant and is used to measure the efficiency of an enzyme, the higher the ratio, the more efficient is the enzyme.

$$E + S \underset{k_{-1}}{\overset{k_1}{\leftrightarrows}} ES \xrightarrow{k_{cat}} EP \leftrightarrow E + P \qquad (2.1)$$

Enzymes discriminate their substrates with great selectivity. An explanation for this was suggested by Emil Fischer who proposed, in 1894, the lock and key theory, according to

which the substrate has to fit into the active site of the enzyme like a key into a lock.[35] This is due to enzymes' complex three-dimensional structures and the active site integrated therein. Additionally, proteins are formed of *L*-amino acids and as a result of this these systems possess a high degree of chemical chirality.[36] Chirality is a property of an object which results in the object being non-superimposable with its mirror image.[37] Two mirror images of a chiral molecule are called enantiomers and a mixture that contains equal amounts of each enantiomer is termed racemic.[37]

Enzymes possess three types of selectivity: chemo-, regio- and enantio- selectivity.[38] Chemoselectivity is the capacity to react specifically with a single type of functional group.[21] Regioselectivity is when two or more identical functional groups are present in the same molecule and the reaction occurs specifically with the functional group that is located in a particular region of the molecule.[21] And finally, enantioselectivity is the capacity to distinguish between different enantiomers of a molecule, either as the reactants or in the generation of the products.[37] Thus, chiral catalysts have the capacity to produce a single enantiomer from a prochiral substrate and to react differently with each enantiomer of a racemic mixture.

The "lock-and-key" model described above considers both ligand and receptors as rigid entities and thus fails to properly represent experimental cases where there is a conformational rearrangement concurrent with the binding event. In order to address the limitations of the "lock-and-key" model, the "induced fit" model was introduced in 1958 by Koshland. This model proposed that the interaction with a ligand can induce a conformational change of the receptor.[39] A third model named "conformational selection"

takes into account the fact that most proteins are dynamic entities and that a vast ensemble of conformational states coexist in equilibrium with different population distributions; a ligand will thus bind to a preferred conformational state and, ultimately shift the equilibrium towards this state.[40-43]

Once the enzymes bind a substrate, the reaction can happen with a catalytic rate constant k_{cat}, which can be linked to the activation free energy of the reaction through transition state theory (TST).[44] **Eq. 2.2** shows the transition state theory formula: where k_B and h are the Boltzmann's and Planck's constants, ξ the reaction coordinate, T the absolute temperature, R is the universal gas constant, K^0 accounts for the standard state, $\Delta G^{\ddagger}$ the free energy difference between the transition state and reactants, γ the transmission coefficient (accounting for the non-separability of the reaction coordinate) and κ the tunnelling contributions.

$$k_{cat} = \kappa\gamma(\xi)\frac{k_B T}{h}K^0 e^{-\frac{\Delta G^{\ddagger}(\xi)}{RT}} \qquad (2.2)$$

From **Eq. 2.2**, the most influential component of the rate constant is the activation free energy ($\Delta G^{\neq}$) because of the rate constant's exponential dependence on this factor. $\Delta G^{\neq}$ represents the free energy difference between the transition state and the reactant state. The transition state (TS) is an intermediate during the conversion of the substrate to the product – a first-order saddle point between the two energy minima.

When reactions are catalysed by enzymes, the $\Delta G^{\neq}$ is significantly decreased, thus increasing the rate constant of the reaction. Consequently, the underlying effect of the enzyme is to change the free energy profile of the catalysed reaction (**Figure 2.3**).

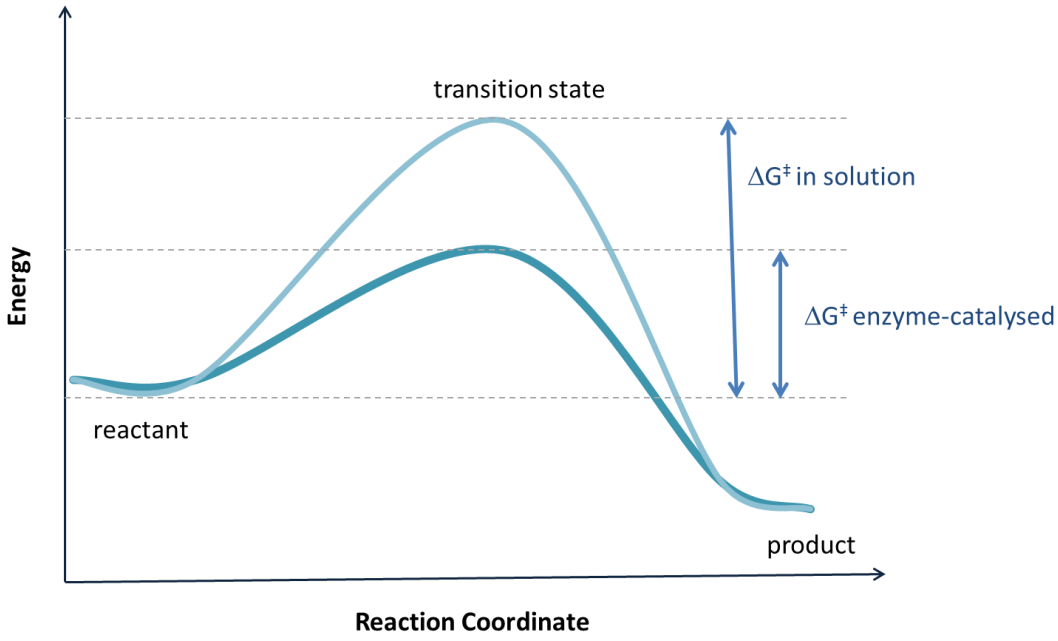


Figure 2.3. Relative free energy profiles for a simple enzyme-catalysed reaction (in dark line) and the uncatalysed process in solution (light line).

As the activation energy is a difference between the two states, a barrier can be decreased either by elevating the energy of the reactant complex or by lowering the energy of the TS.[45] The reactant state can be destabilised by the enzyme applying a structural constraint on the substrate to make its structure closer to that of the TS.[45-46] In addition, the TS can be stabilised by additional interactions that will lead to a decrease in the activation energy.[47-48]

Theoretical studies using methods and techniques of computational chemistry have helped in the understanding of the physical basis of rate enhancement of chemical reactions by enzymes.[45, 49] Warshel and co-workers have compared the contribution of several catalytic factors on a wide range on enzymes and concluded that electrostatics is the key catalytic factor.[50] Work from other groups have also supported this proposal.[51-52] The importance of electrostatics is explained by the pre-organized environment that the enzyme active site offers to the reaction. Indeed, the active site has been shown to provide an electric field prepared to accommodate the charge distribution of the transition state without changing the enzyme environment too much.[53-55] In contrast, in aqueous solution, water molecules have to spontaneously order around the reacting species, which has an energetic cost (**Figure 2.4**).

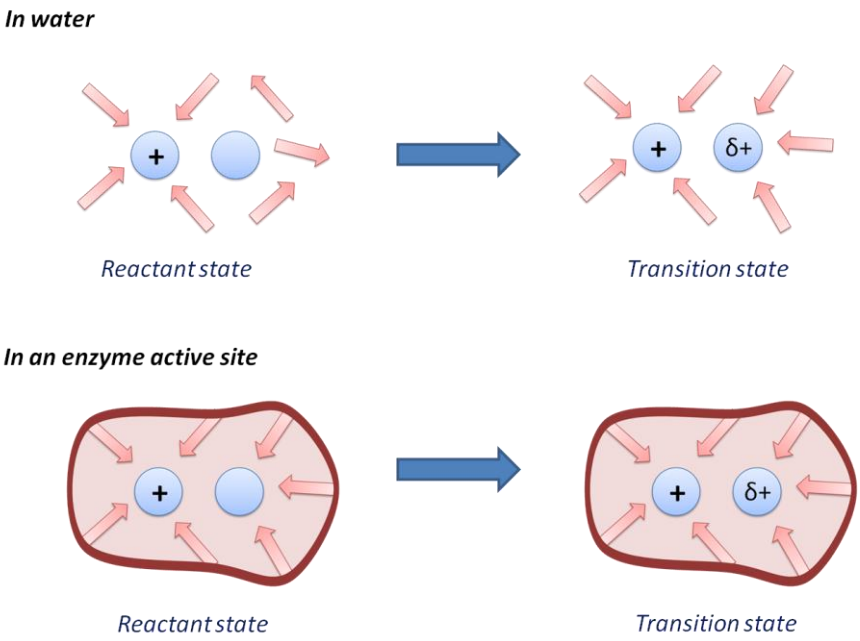


Figure 2.4. Contribution of the pre-organized environment to the transition state stabilisation.

More recently experimental studies have confirmed the catalytic role of external electric fields in enzymes.[56] Boxer's group has provided strong evidence for this concept on the basis of the vibrational Stark effect (VSE).[57] Vibrational Stark spectroscopy has been used to measure local dipole moments in the active site that are influenced by an external field in order to quantify their contribution to the observed acceleration of reaction rates over the uncatalysed reaction in aqueous solvent. For Ketosteroid isomerase (KSI) Boxer and co-workers have shown that the large electric fields exerted on the active site were linearly correlated with the activation free energies of the wild type and mutated variants.[56]

In recent years, there have been some debates on whether dynamic effects should complement our understanding of enzyme catalysis.[58-62] There is today a consensus on the importance of flexibility and dynamic effects on the catalytic process (*i.e.*, for ligand binding and product release). However, the current debate is about whether enzymes 'dynamics' contribute significantly to the rate enhancement achieved by an enzyme. Part of the debate has been linked to different definitions of 'dynamics' and 'catalysis'.[62] Enzymes dynamics can refer to wide-ranging timescales (from seconds to femtoseconds) and thus to different amplitudes of motion (from large domain motion to bond vibration).[63] Also 'dynamics' can refer either to statistical dynamics (from thermal equilibrium following the Boltzmann distribution) or nonstatistical dynamics.[62] The definition of enzyme catalysis refers to the rate enhancement achieved by an enzyme relative to a uncatalyzed reaction without enzyme (k_{cat}/k_{uncat}) and is thus only associated with the chemical step of the catalysis cascade (which also include substrate binding, structural rearrangement or product release).[64]

Part of the debate originated from the experimental observation of an unusual temperature dependence of kinetic isotope effects (KIEs).[65] A KIE can happen during a hydrogen transfer reaction and refers to the significant change in the measured reaction rate constant observed when using heavier isotopes of hydrogen (*e.g.*, deuterium). Until recently TST was not able to reproduce some temperature dependencies in these KIE.[65] These observations have motivated the development of new models that explained these phenomena as the consequence of protein motions.[65] One of these models is 'promoting vibrations'[66] and refers to fast (femtosecond to picosecond) dynamics involved in transition state barrier crossing.[67-68] Another model is 'promoting motions' and suggests the existence of equilibrium thermal motions on the millisecond timescale that promote an active site organisation favourable to catalysis.[69-70] However, In 2012 Glowacki *et al.*, provided a simple model using TST that could properly explain the previously observed unusual temperature dependence of KIEs; the trick was to incorporate conformational sampling into the TST by using at least two representatives of the protein conformation for the calculations.[71] Their result shows that no new theoretical framework needs to be invoked. Additionally, the use of TST allows remarkable agreement with experiment which is an argument that this framework is sufficient to explain catalysis even if it does not refute the dynamic hypothesis.[72] However the direct link between enzyme dynamics and chemical reaction rate (or energy barrier lowering) has been carefully examined by computer simulations and the resulting conclusion is that it is likely that the contribution of such effects are small;[73-74] the major contribution to the increased rate observed in enzymes seem to arise from the electrostatic stabilisation of the transition state.

2.2 Biocatalysis for pharmaceutical manufacturing

Biocatalysis is the use of enzymes in chemical synthesis. Enzymes used for biocatalysis can be obtained from natural sources or produced using recombinant DNA, the latter method allows for the possibility of more efficient, higher purity and cheaper production. They can then be used in whole cells, lysates or in purified forms. Biocatalysts can be used as the free enzyme or bound to a solid support (immobilisation). Immobilisation has several advantages such as a potential increase in activity and/or stability compared to the soluble equivalent.[75-76] Also, it allows enzyme reuse which can considerably decrease the cost of the process. Examples of commonly used immobilisation techniques are covalent attachment, adsorption, crosslinking or encapsulation.[75] Biocatalysis reactions can take place in water but to expand the set of possible substrates that can be used, organic solvents would be preferred. Organic synthesis often deals with organic substrates, which are poorly, or not, water soluble. These solubility problems can be overcome by switching from water to organic solvents as the reaction media. In 1984, the seminal work of Klibanov demonstrates that enzymes could be use in organic solvent with similar activity to that obtained in water, which opened the door to a much broader scope of applications.[77] More recently, Ionic liquids (IL) are increasingly being used as greener alternatives to organic solvent, which have the inconvenience of being volatile and potentially flammable.[78] IL are salts in liquid form with negligible vapor pressure which make them non-flammable and highly thermostable. Additionally, IL has the advantage of being able to dissolve both polar and nonpolar substrates.

Biocatalysis has been increasingly used for the synthesis of pharmaceuticals.[79] The biocatalytic toolbox is continuously expanding and includes representatives from across almost all the EC enzyme classes (**Table 2.2**). Oxidoreductases, transferases, hydrolases and lyases have been broadly and successfully used for the biocatalysis of pharmaceutical intermediates. Application of isomerases and ligases are still limited but recent developments could make them emerge as interesting biocatalysts for new synthetic applications.[80-81]

Table 2.2. Key enzyme classes applied for the synthesis of pharmaceuticals.

Enzyme class	**Biocatalyst**	**Example of application**
Oxidoreductase	Ketoreductase (EC 1.1.1.2)	Atorvastatin (Lipitor®)[82]
	Monoamine oxidase (EC 1.4.3.4)	Boceprevir (Victrelis®)[83]
Transferase	Transaminase (EC 2.6.1)	Sitagliptin (Januvia®)[84]
Hydrolase	Lipase (EC 3.1.1.3)	Ibuprofen[85]
	Penicillin acylase (EC 3.5.1.11)	Amoxicillin[86]
Lyase	Aldolase (EC 4.1.3.3)	Zanamivir (Relenza®)[87]

2.2.1 Biocatalysis as a means to sustainable chemistry

In the pharmaceutical industry, the development of biocatalysis has in part been driven by the need for more sustainable manufacturing process. In 1992 a metric for assessing the environmental impact of manufacturing processes named the E(environmental) factor was

introduced by Roger Sheldon.[88] It refers to the mass of waste per mass of product and is usually expressed as kilograms per kilogram [kg/kg]. The lower the Efactor the better and the target must be the ideal value of zero. The E factor of various industries was calculated and it was found that the pharmaceutical industry had the highest result.[88] The quantification of the waste generated has certainly promoted the need for change. In the early 1990s the term "green chemistry" also emerged at the U.S. Environmental Protection Agency (EPA). The 12 green chemistry principles give a set of rules to follow in order to have more sustainable processes.[89]

Biocatalysis fulfils most of the 12 principles and has thus become a key technique to perform sustainable chemistry.[6] First of all biocatalysts come from renewable sources and are biodegradable. Also, enzyme-catalysed reactions can often be carried out at ambient temperature and atmospheric pressure, thus avoiding the use of more extreme conditions which can cause problems with isomerisation, racemisation, epimerization, and rearrangement.[79] Additionally, their high selectivity make them more economic industrial processes as multiple process steps can be circumvented.[90-91] The use of biocatalysts can reduce or eliminate the requirement for protective group chemistry, thus the number of steps is reduced, and each step that is removed reduces the requirement for reagents, solvent, energy and time decrease the amount and toxicity of waste products.[92]

One example is the production of penicillin-based semi-synthetic antibiotics. Penicillin was discovered in 1928 by Alexander Fleming.[93-94] In 1957, the discovery of 6-aminopenicillanic acid (6-APA), the so-called penicillin nucleus, allowed the production of many semi-synthetic penicillins.[95] The discovery of these semi-synthetic penicillins

improved the range of infections that could be treated by penicillins. Originally the production of semi-synthetic antibiotics required the use of organic solvents as well as the reactive group's protection/deprotection sequence, low temperatures (-30°C), and chemical acylation. In the biocatalytic processes, the chemical removal of side chains is replaced by an enzymatic reaction step (hydrolysis) using immobilized penicillin acylase (EC 3.5.1.11) that is regio- and stereo-specific and the reaction conditions are mild (**Figure 2.5**).[86] Interestingly, the same enzyme also catalyzes the reverse reaction so it allows for the coupling of different side chains. Amoxicillin (D-(-)-α-amino-*p*-hydroxybenzyl penicillin trihydrate) is an example of a semi-synthetic antibiotic currently produced by this strategy.[86, 96] This drug has a broad spectrum of bactericidal activity against gram-positive and gram-negative microorganisms and is one of the major β-lactam antibiotics, which, according to IMS Health, was part of the top 15 dispensed prescriptions in 2012 in the United States of America.[97] The production of semi-synthetic β-lactam antibiotics is the highest volume application of biocatalysis within GSK with over 1,000,000 kg annually.[98]

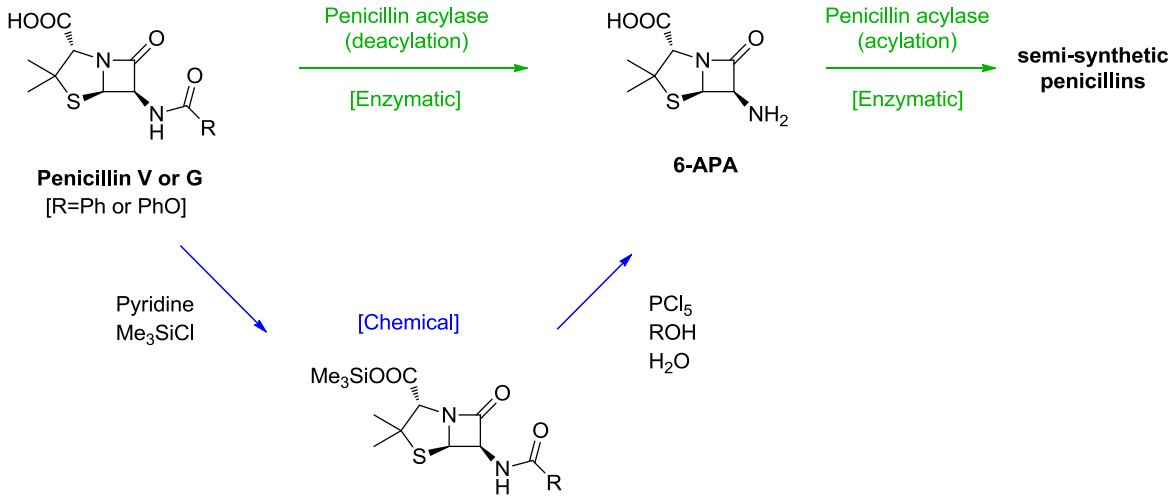


Figure 2.5. Chemical and enzymatic deacylation of penicillins to 6-APA.

2.2.2 Biocatalysis for the synthesis of enantiopure compounds

Another factor that has driven the increased use of biocatalysis in the pharmaceutical industry is the increasing demand for single-enantiomer drugs (**Figure 2.6**).[99-100] Indeed the ability of different enantiomers to interact selectively with biological receptors is an important property, which can result in a profound effect on drug–receptor interactions.[101] That is, two enantiomers of the same drug may not have the same effect and sometimes need to be considered as two different drugs.[101]

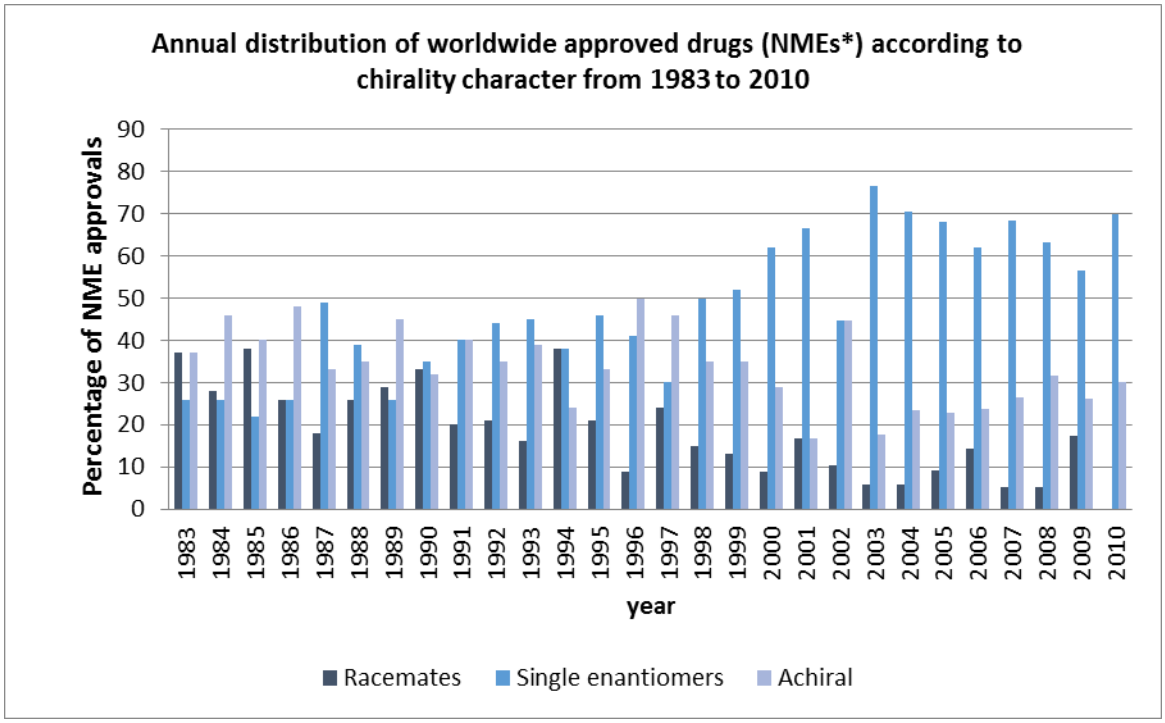


Figure 2.6. Annual distribution of worldwide-approved drugs according to chirality in the period 1983–2010. Data collected from [102-103] *According to the US Food and Drug Administration, a New Molecular Entities (NME) is an active ingredient that has never been marketed in the US before in any form; it is a term which is generally recognized world-wide.[104]

An illustration of the importance of drug stereochemistry is provided by the well-known example of the thalidomide disaster. Thalidomide was widely prescribed for morning sickness from 1957 to 1962.[105] However, despite its success with curing morning sickness, the drug was subsequently shown to be associated with serious teratogenic effects.[106] Approximately 10,000 children were born with thalidomide-related disabilities worldwide.[105] The drug was administrated as a racemic mixture but there is some evidence that whereas the (*R*) isomer may be an effective sedative, the opposite (*S*) isomer would

be associated with teratogenic properties and could be responsible for the appearance of deformities in the developing foetus.[107] The differential effect of both enantiomers remains hypothetical, indeed the individual effect of each enantiomer is difficult to evaluate mainly because a chiral inversion can occur *in vivo*.[108]

Nevertheless, the thalidomide disaster led to a change in drug testing and approval procedures in many countries, such as the Kefauver-Harris Amendment (U.S.) and Directive 65/65/EEC1 (E.U.).[109-110] The first of these laws required drug manufacturers to provide proof of the effectiveness and safety of their drugs before approval and the second was the first European pharmaceutical directive and required that a proprietary medicinal product could not be marketed within the community without prior authorisation of the competent authority of at least one member state.

Another example of a drug which has differential effects between its enantiomers is ibuprofen.[111] This well-known anti-inflammatory and analgesic agent has been in use for over 40 years, and is also administered as a racemic mixture.[111-112] However, it has been revealed that only the (*S*) isomer is a cyclooxygenase inhibitor and the origin of the anti-inflammatory effect.[111]

Enantiopure drugs also have an economic value as pharmaceutical companies can use chirality as a tool to increase the span of their patented blockbuster drugs.[113] Indeed, "chiral switches" are chiral drugs that have already been claimed, approved and marketed as racemates or as mixtures of stereoisomers, but have since been redeveloped as single enantiomers.[99, 113] According to the US Food and Drug Administration (FDA), single

enantiomers in chiral switches are not new molecular entities, and are therefore barred from five-years exclusivity.[113] Such new products are treated as new derivatives of existing drugs or new formulations, on a case-by-case basis, and can be eligible for three years exclusivity.[113]

Given their important properties and economic value, the development of new processes for the production of chiral molecules remains an ongoing challenge in organic synthesis.[114-116] Currently the two main strategies for the synthesis of single-isomer drugs are isomer separation and asymmetric synthesis (**Figure 2.7**).[115, 117] The "isomer separation" strategy involves starting from a racemic solution and applying a method, such as chiral column chromatography, resolution by crystallization, use of chiral resolving agents or enzymatic resolution,[117] to separate the enantiomers.

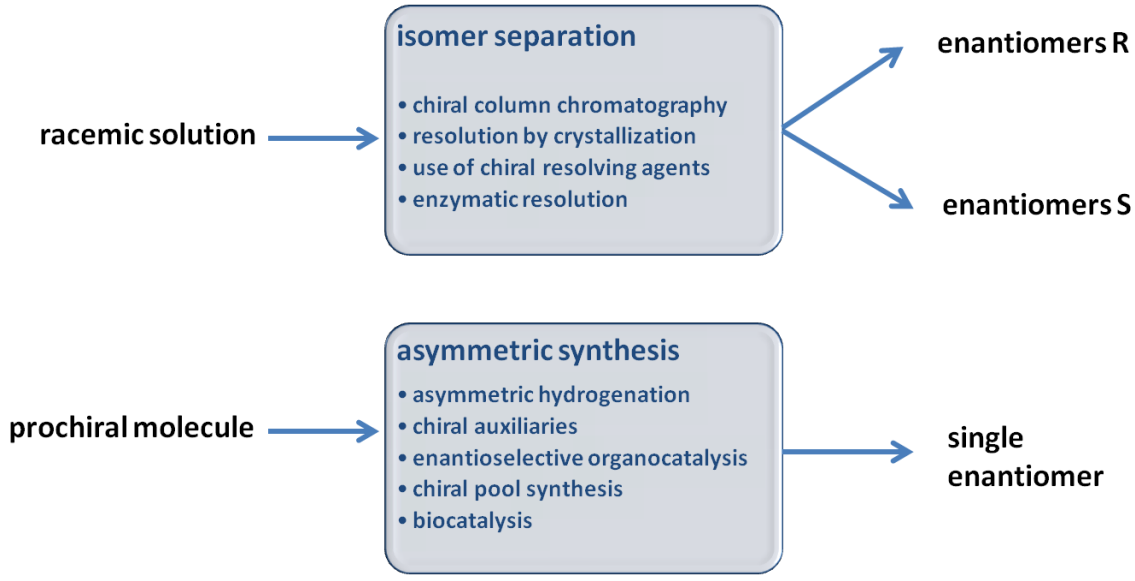


Figure 2.7. Approaches to the isolation of enantiopure compounds. Information collected from [115, 117]

When only one enantiomer is needed another strategy is to perform asymmetric synthesis using enantioselective catalysis. There are several approaches that can be used to achieve this, such as asymmetric hydrogenation,[118] chiral auxiliaries,[119] (an organic compound which couples to the starting material to form new compound, which can then undergo enantioselective reactions via intermolecular asymmetric induction), enantioselective organocatalysis,[120] chiral pool synthesis[117] (using chiral natural compounds as starting point) and biocatalysis.[116] The use of biocatalysts for chiral synthesis is currently preferred because it offers several advantages including very high enantiomeric excess (ee, the degree to which a sample contains one enantiomer in greater amounts than the other) and reagent specificity, as well as mild operating conditions and low environmental impact.

Lipases (EC 3.1.1.3) are one of the most used enzymes in the synthesis of enantiomerically pure intermediates.[121] They have the advantage of being stable in organic solvents and are cofactor-independent, along with high activity and broad substrate acceptance.[122] Lipases can catalyse the synthesis of enantiomerically pure intermediates via hydrolysis or transesterification.[121] One example of lipase application is the separation of the racemic solution of ibuprofen and other types of 2-aryl-propionic acids derivatives through enzymatic resolution.[123] Both esterification and hydrolysis approaches with lipases have been successfully applied to the production of (*S*)-ibuprofen (**Figure 2.8**).[124-126] In these systems, the lipase selectively generates one enantiomer – *e.g.*, CRL selectively reacts with (*R*)-ibuprofen to generate the (*R*)-ibuprofen, *n*-propyl ester, leaving the (*S*)-ibuprofen unreacted (**Figure 2.8a**) – and the resulting compounds can be readily separated due to

their distinct physiochemical properties. Similar strategies have been for example applied to the production of the chiral intermediate of the anti-hypertension drug diltiazem.[127-128]

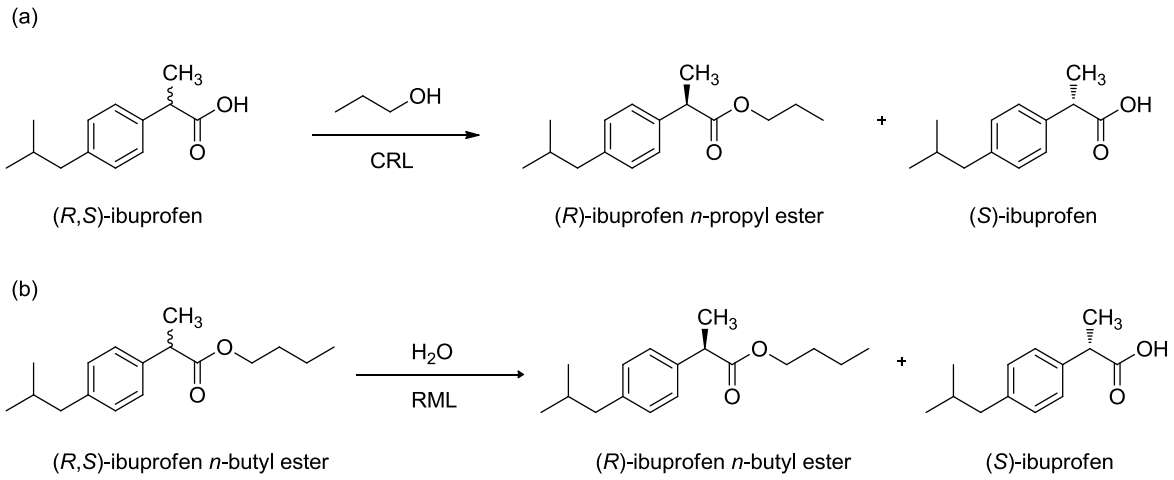


Figure 2.8. Lipase-catalysed enantioselective (a) esterification and (b) hydrolysis of racemic mixture. CRL: *Candida rugose* lipase; RML: *Rhizomucor miehei* lipase

Enzymatic resolution and other separation techniques have the intrinsic disadvantage of producing a maximum theoretical yield of 50 %. To overcome this limitation another strategy is to employ enzymes capable of performing asymmetric synthesis on prochiral molecules. Ketoreductases (KREDs, EC 1.1.1.2) catalyse the NADPH dependent asymmetric reduction of ketones to chiral alcohols. They have for example found application in the synthetic route of blockbusters such as the lipid-lowering agent atorvastatin (Lipitor®) and the asthma maintenance treatment montelukast (Singulair®, **Figure 2.9**).[82, 129]

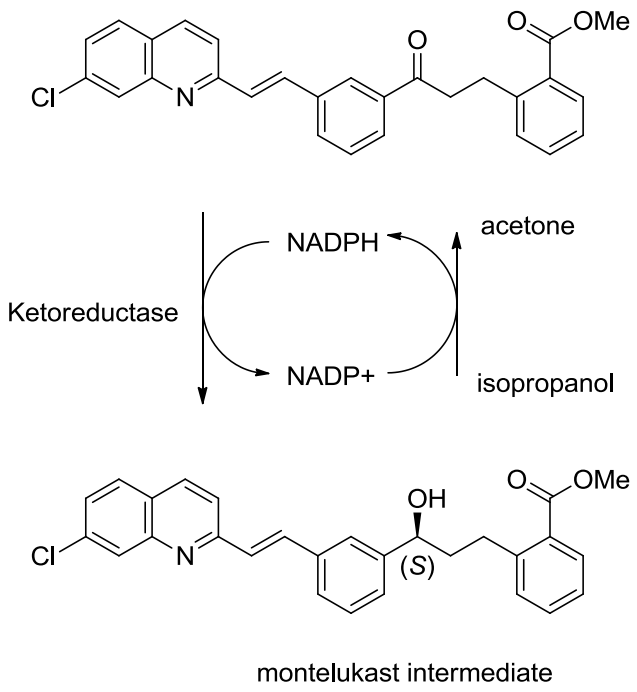


Figure 2.9. Ketoreductase-mediated synthesis of a montelukast intermediate.

Reductive amination methods for production of chiral amines are particularly important as it is estimated that 40 % of pharmaceuticals contain a chiral amine moiety.[130] Various biocatalytic strategies have been developed for that goal. Here again lipases have been used via enzymatic resolution of racemic amines.[131] During the last decade transaminases have matured to a robust method for the preparation of optically pure amines.[132] Transaminases (TAms, EC 2.6.1) are enzymes that catalyse the pyridoxal-5'-phosphate (PLP) dependent transfer of an amine group from an amine donor to a ketone or aldehyde. One famous example of integration of a transaminase is the development of a new process for sitagliptin (**Figure 2.10**).[84] Sitagliptin is the active ingredient in Januvia®, a treatment for type 2 diabetes and is the largest product by both volume and sales of Merck.[133] The original process involved an asymmetric hydrogenation of an enamine at high pressure

using a rhodium-based chiral catalyst for which researchers from Codexis and Merck developed a biocatalytic alternative.[134] The study started with a (*R*)-selective transaminase ATA-117 from *Arthrobacter* sp. selected for asymmetric amination of pro-sitagliptin. The initial enzyme did not show any activity for the substrate due to steric constraints. *In silico* studies were carried out to investigate key residues to mutate to fit the substrate in the pocket which combined with directed evolution could produce the enzyme needed for commercial level. The protein engineering procedure took 11 rounds of directed evolution and 27 random mutations were introduced. The obtained variant is able to convert 200 g/L of the pro-sitagliptin ketone into sitagliptin in 50% DMSO with 92% yield (>99.95% ee). Relative to the original chemical route, this enzymatic process generates a 10% increase in overall yield and 53% increase of productivity (kg/L per day). Importantly this also reduce the total waste by 19%. This process has been awarded the prestigious Presidential Green Chemistry Challenge Awards. Transaminases have recently been applied on different pharmaceutical production in both enzymatic resolution and asymmetric synthesis of the ovarian cancer treatment niraparib,[135] the anti-insomnia drug candidate filorexant (MK-6096),[136] and the anti-tumor drug candidate AZD1480.[137]

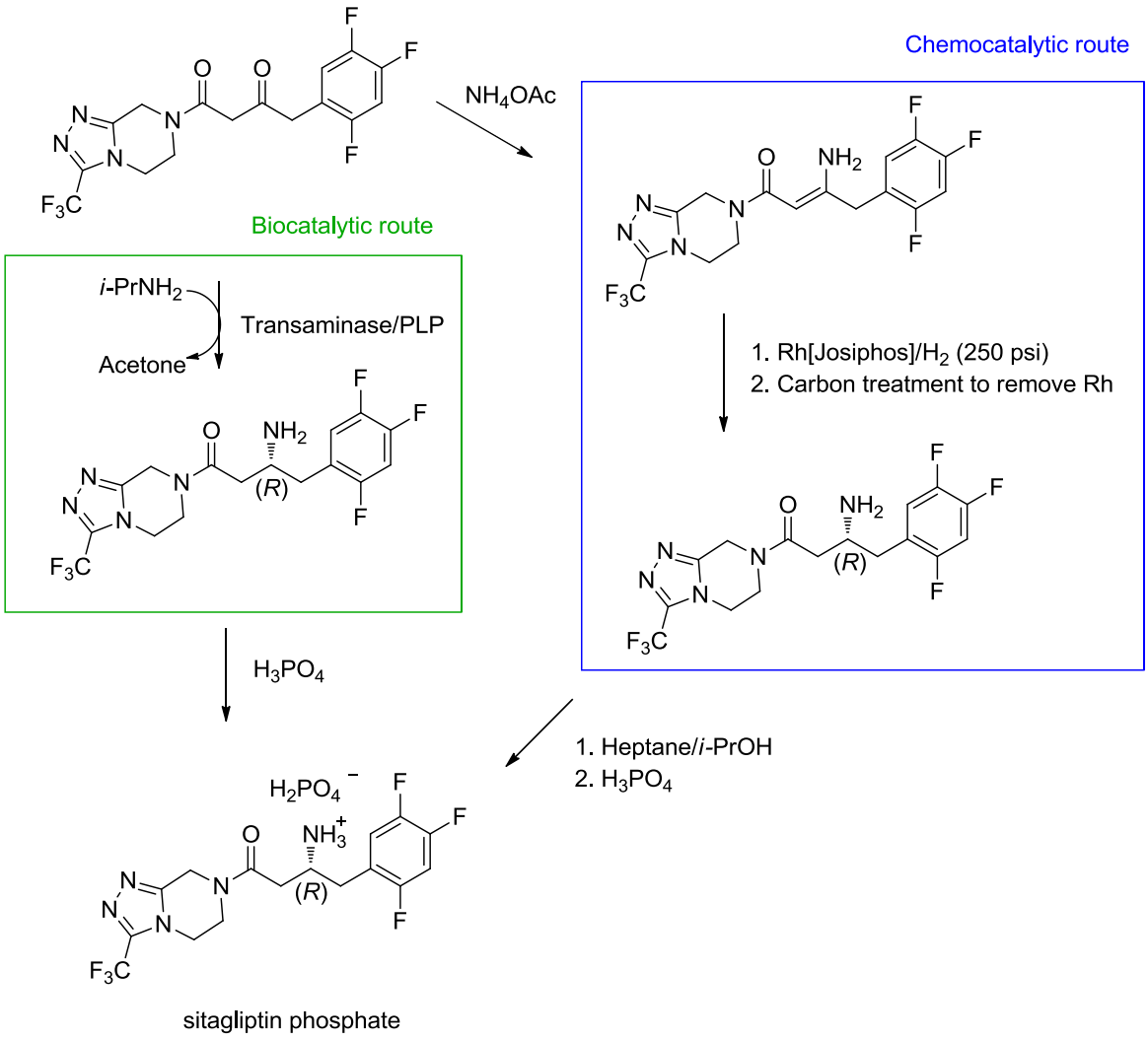


Figure 2.10. Biocatalytic and chemo catalytic routes for the asymmetric synthesis of sitagliptin.

Lipases and transaminases provide useful tools for the production of chiral primary amines but robust biocatalyst platforms for secondary and tertiary amines are still in development. Different oxidoreductase enzymes have been explored for these kinds of applications which include, among others, Monoamine Oxidases and Imine Reductases.[138] The advance in available regeneration methods for redox cofactors has allowed the broader

application of this class of enzyme to industrial scale.[139] Monoamine Oxidases (MAOs, EC 1.4.3.4) are flavin-dependent enzymes that catalyse the oxidation of amines to imines (**Figure 2.11a**). The capabilities of MAOs for the production of enantiopure amines on various substrates was extensively developed by Turner's group.[140] When coupled with nonselective chemical reducing agents MAOs can mediate the deracemization of simple chiral primary, secondary, and tertiary amines. Industrial application of this process for the production of the Hepatitis C drug boceprevir (Victrelis®) was recently developed.[83] Imine Reductases (IREDs, EC 1.5.1.48) catalyse the NADPH-dependent enantioselective reduction of imines to the corresponding amine (**Figure 2.11b**).[141] Two particularly interesting strains of IREDs were discovered by Mitsukura and co-workers around 2010.[142-143] The current challenge consists in developing broader substrate scope to allow further synthetic applications.[144]

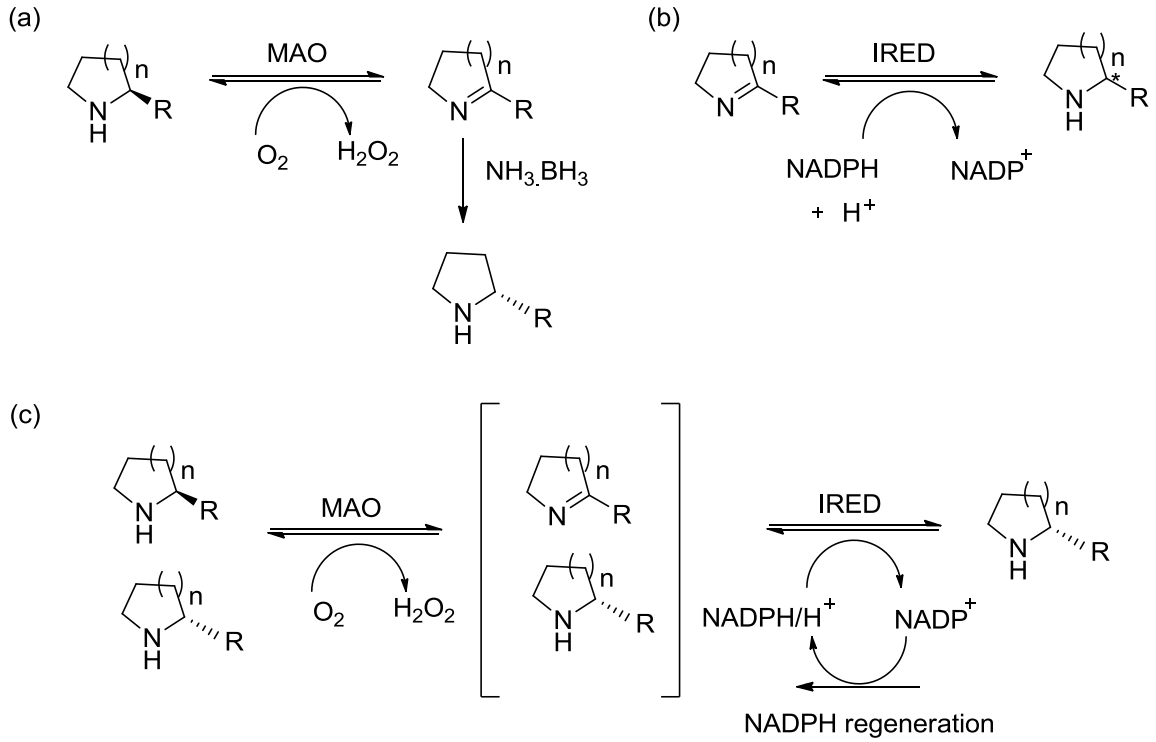


Figure 2.11. Transformation catalysed by (a) monoamine oxidases (MAOs), (b) imine reductases (IREDs). (c) Deracemization catalysed by a MAO/IRED cascade.

The increased availability of enzymes motivates the increased development of multi-enzyme systems to perform cascade reactions.[145] For example IREDs have been applied as the reducing agent in MAO-catalysed deracemisation cycles leading to enhanced enantioselectivity in the production of piperidine and pyrrolidine substrates (**Figure 2.11c**).[146] Cascade reactions have the advantage of avoiding elimination and purification steps and thus result in reduced cost and waste. These approaches can also be used for by-product removal. There are numerous reports of biocatalytic cascades particularly involving TAms.[147] Another on-going development is the engineering of metabolic pathways for the production of pharmaceutical intermediates. However, these are still

mostly limited to the production of bulk chemicals. One example is the biosynthesis of the non-natural amino acid *L*-homoalanine,[148] which is a chiral precursor for the tuberculosis medications ethambutol (Myambutol®) and the anti-epileptics drugs levetiracetam (Keppra®) and brivaracetam (Briviact®).

2.2.3 Protein engineering of biocatalysts

The industrial application of penicillin acylases for the production of 6-APA was greatly facilitated by the fact that the wild-type enzyme identified by screening already had satisfactory yields and no protein engineering was required.[149] However in the majority of cases, naturally occurring enzymes are often not suitable for biocatalytic processes without further tailoring or redesign of the enzyme, due to limitations regarding thermostability, activity, selectivity (narrow range of natural substrates), and tolerance towards organic solvents for industrial applications.[150] Although process engineering can sometimes solve some of these problems, for example through enzyme immobilisation, it is usually not sufficient.[76, 151] Therefore, protein engineering strategies aimed at constructing enzymes with novel or improved activities, specificities, and stabilities have been developed.[152]

One strategy in protein design is the design of proteins from scratch, known as *de novo* design.[153] Last decade has seen the appearance of the first *de novo* design of enzymatic activity that was not reported before, a Kemp-elimination activity.[154] *De novo* design of enzymes with Retro-aldolase activity and stereoselective Diels-Alder activities have also been reported.[155-156] These designs were based on the Baker lab algorithm,[157] and follow

the "inside-out" strategy.[158] This strategy follows two steps : first the design of an optimal active site to accommodate the reaction (named a "theozyme" and obtained from quantum mechanics calculations), second the identification of a scaffold using the RosettaMatch[157] module and third the mutation and optimization of the amino acid residues surrounding the "theozyme". Activities of these designs are low but it was found that by further refinement through directed evolution higher rates could be obtained.[159-160] Nevertheless these computationally designed enzymes perform quite poorly and have not found industrial application so far.

An alternative to *de novo* design is the redesign of existing enzymes. Key advances in DNA sequencing and gene synthesis are at the base of tremendous progress in tailoring biocatalysts by protein engineering and design.[161] Strategies for protein engineering can be classified as random, rational and combined methods (semi-rational).

2.2.3.1 Random approaches

Directed evolution (DE) has proved to be an effective strategy for improving biocatalyst properties starting from wild-type enzymes.[162] Specific activity, stability, substrate scope, and stereoselectivity of enzymes can be optimized using this technique. This technique consists of subjecting a gene to iterative rounds of variation (creating a library of variants), selection (expressing the variants and isolating members with the desired function), and heredity (generating a template for the next round) (**Figure 2.12**).

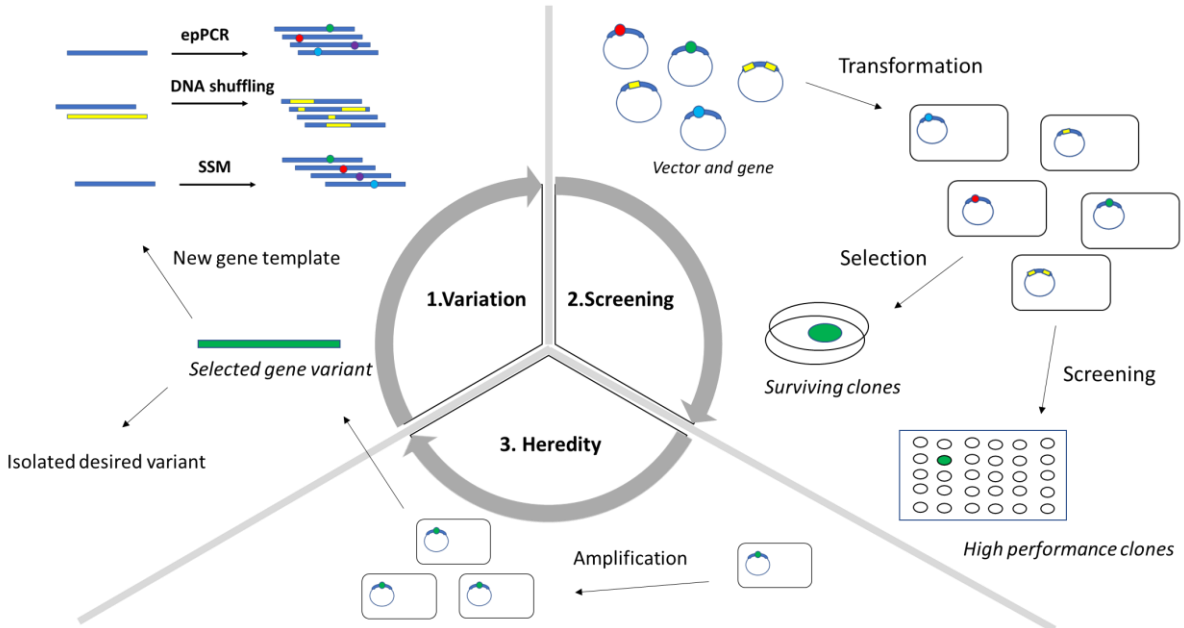


Figure 2.12. Schematic representation of a directed evolution process.

Established methods for the library generation in DE are error-prone PCR, DNA shuffling and site-saturation mutagenesis (step 1 of **Figure 2.12**). Error-prone Polymerase Chain Reaction (epPCR) was first introduced in 1989 by Leung.[163] By using the low-fidelity DNA polymerase (*i.e.*, *Thermus aquaticus*) it allows random introduction of error in the sequence. The DNA shuffling method is for *in vitro* recombination of homologous genes and was developed by Stemmer in 1994.[164] A set of naturally occurring homologous sequences are randomly fragmented using a DNase and are then recombined to form chimeric entities. Site-saturation mutagenesis (SSM) is a method that allows the systematic substitution to all of the other 19 possible amino acid at a particular position.

After the variation step these genes are cloned into DNA vectors and transformed into an expression host for expression of the corresponding protein variants (step 2 **Figure 2.12**).

It is worth noting that the transformation efficiencies are far from optimal and that a lot of the variants that are generated in the first step will not be transformed. Following the transformation step the protein variants need to be evaluated using agar plate or microtiter plate-based screening. Agar plate-based screening allows better throughput capacity (10^5 variants analyses per day) but cases where it can be used are limited.[165] Therefore, a majority of library analyses continues to be performed via screening in microtiter plates instead. Microtiter plate screening is usually limited to no more than 10^4 variants analyses per day.[166]

In the last step of a DE a desired variant is selected and amplified to be used as a template for the next round of directed evolution (step 3 of **Figure 2.12**). The amplification is done by bacterial growth of cells from microtiter plates' wells with the highest detected activities or of the surviving colonies from agar plates. The cycle in **Figure 2.12** will be repeated until a variant with the desired property is obtained.

In the area of DE, new opportunities are arising from the development of ultrahigh-throughput screening/selection technologies that allow more than 10^5 analyses per day. These technologies allow one to increase the throughput of screening by replacing the classic microtiter plates by smaller systems (*e.g.*, droplets). These compartments can be analysed by fluorescence-activated cell sorting (FACS).[167-168] Flow cytometers enable screening of up to 10^8 variants per day. Desired activities can thus be reached with fewer cycles.[169-170] These technologies are interesting because they allow escaping from microtiter plate-based DE plateaus,[170-171] the analysis of metagenomics libraries and broadening the applicability of deep mutational scanning.[172] Deep mutational scanning is

the principle of systematic analysis of the DE outcome to understand it better.[173] This approach can now be done by combining ultrahigh-throughput with next-generation sequencing. The motivation for this kind of study is that a better understanding of the structure-sequence relation would allow one to accelerate the enzyme optimisation process. These studies typically focus on the consequences of single mutations.[174] However, a more complex analysis can investigate the effect of multiple mutations and give insight into epistatic interactions, which are mutation that act additively.[175]

Despite the impressive advances made in these fields, the challenge of specifically designing biocatalysts for a given function or substrate still remains a significant hurdle.[176] The size of these libraries is often a bottleneck of this method, particularly for the selection parts.[177] Large libraries are actually needed because, as they are usually built randomly, the rate of success is usually low.[178]

2.2.3.2 Semi-rational approaches

Semi-rational design randomizes only specific amino acids within the enzyme's protein sequence that are deemed to be of interest. The advantage of semi-rational methods is efficiency, as the same or better results can be obtained with less effort.[179-180] Semi-rational strategies can be done for example to optimize substrate specificity and enantioselectivity as these features are often governed by steric factors of the active site. Iterative Saturation Mutagenesis (ISM) is a well-established method for semi-rational design.[181-182] In this method multi-residue sites of interest are first identified. In a second step, residues in each multisite are simultaneously subjected to saturation mutagenesis. Then the chosen variant from one multisite is used as a template for saturation mutagenesis

at other multisite. The ISM method typically used to improve selectivity or activity is the Combinatorial Active-Site-Saturation Test (CAST).[183] To improve stability another ISM method named B-Factor Iterative Test (B-FIT) can be used in which only residues with the highest crystallographic B-factor (more flexible) are mutated.[184]

Awareness of the benefit of this type of technique was gained by the example of *Pseudomonas aeruginosa* lipase (PAL) for which it was shown that by using an ISM method the desired enantioselectivity could be reached faster than by epPCR (10,000 versus 50,000 transformants).[185] epPCR or DNA shuffling should thus be used only if no structure based information is available. Different variants of ISM have been recently developed such as single code saturation mutagenesis (SCSM)[186] or triple code saturation mutagenesis (TCSM).[187]

2.2.3.3 Computational methods for Hotspot identification

One option to narrow and improve the quality of variants' libraries can be the use of computational methods.[188] Computational methods can be applied to biocatalyst design in order to gain rational guidelines, to orient experimental planning and, ultimately, to avoid expensive and time-consuming experiments.[189]

Knowledge-based approaches rely on experimental and sequence data for guidance. Some widely used software are Hot-Spot Wizard[190], ProSAR[191] and SCHEMA[192]. Platforms that contain information on enzyme superfamilies' superimposition have been developed which can be useful to provide sequence-function relationships with the well-known example of the 3DM commercial database.[193-194] ZEBRA is a web server for analysing

enzyme functional subfamilies in order to identify adaptive mutations.[195] A test case was run on the *Candida antarctica* lipase B (CALB) by integrating subfamily specific positions calculated for the amidases which resulted in an improved amidase activity of CALB.[196]

Physics-based approaches utilize principles from computational chemistry to predict properties of proteins only from their structure.[197] Structural biology techniques such as protein crystallography or NMR spectrometry allow the determination of protein structures to atomic resolution and these can be employed as starting points for molecular modelling studies. Physics-based methods usually focus on the improvement of one specific property which can be the protein stability, the protein-ligand binding or the catalytic rate constant.

For stability prediction, different methods that estimate the folding free energy are available such as Rosetta[198-199] and FoldX[200-201]. Recently, the melting apparent temperature (T_m) value of a transaminase was increased by 4.0°C relative to the T_m of the wild-type enzyme by using FoldX.[202] A more recent approach developed by the Janssen group is named FRESCO (Framework for Rapid Enzyme Stabilization by COmputational libraries) and runs a MD-based screening on a library of single point mutations generated from Rosetta, FoldX and an in-house code for disulfide discovery.[203-204] By employing this FRESCO protocol on a limonene epoxide hydrolase (EC 3.3.2.8) from *Rhodococcus erythropolis* the T_m was increased from 50 to 85°C.[203]

To improve selectivity, it is the binding free energy of the protein-ligand complex that is estimated by programs such as Rosetta.[205] The selectivity improvement protocol usually includes the creation of *in silico* variants, the optimisation of their structure (via sampling) and their evaluation by using an energy function. Here again the Janssen group has recently developed a MD-based protocol name CASCO (CAtalytic Selectivity by COmputational design) which integrates RosettaDesign in a first step for variants generation and allows stereoselective enzyme design.[206] By employing the CASCO protocol, an epoxide hydrolases that produced diols with high enantiomeric excess was discovered.[206] Another recent advance in that area is the development of programs allowing backbone flexibility such as RossetaRemodel or OSPREY.[207-208]

The accurate prediction of catalytic rate constants needs the identification of the transition state (TS) of the reactions to evaluate the reaction energy barrier. However, the identification of transition states is usually a time-consuming task. Two main strategies have been used for a fast estimation the catalytic rate constant: (a) simplify the activation energy barrier (E_a) calculation procedure; or (b) evaluate the transition state stabilisation by estimating the enzyme-transition state binding energy (ΔG_{bind}(E-TS), **Table 2.3**). Another important parameter is whether the hotspots are identified using quantum mechanics or molecular mechanics. The first allows the attainment of more accurate results while the later allows more sampling.

Table 2.3. Selected examples of strategies to estimate catalytic rate constants.

Study	Property estimated	Energy	Sampling
Hediger *et al.*[209]	E_a	PM6	No
Funke *et al.*[210]	E_a	BLYP/6-31G*: CHARMM	No
Zheng *et al.*[211]	ΔG_{bind}(E-TS)	AMBER	MD
	E_a	B3LYP/6-31G*: AMBER	
Gu *et al.*[212]	ΔG_{bind}(E-TS)	AMBER	MD
OptZyme[213]	ΔG_{bind}(E-TS)	CHARMM	MD

To simplify the energy barrier calculation Jensen and co-workers have investigated the use of semi-empirical methods (PM6) rather than the more accurate *ab initio* methods (*e.g.* B3LYP) to identify beneficial mutations.[209] Their protocol, applied to CALB, has allowed the identification of improved variants for the amidase activity.[214] A different strategy was taken by Funke and co-workers,[210] in their study they set side chain atomic point charges of single residues to zero and re-evaluated the activation energies by simple single point calculations on both the reactant and the transition state in the modified electrostatic environment. This protocol is motivated by the idea that electrostatics is believed to be one of the most important components in determining the catalytic ability of enzymes.[50] The methodology was applied to *Bacillus subtilis* lipase A and the results identified one

hotspot (ΔE_a < -1 kcal/mol) in His76 with a ΔE_a of -2.1 kcal/mol. Parallel experimental work also identified His76 as an important residue for catalytic activity.

Methods based on enzyme-transition state interaction energy estimation have been employed by Zheng *et al.* to design a cocaine hydrolase (EC 3.1.1.84).[211] In their study the TS of the rate-determining limiting step was first determined for the wild-type structure. Then the enzyme-TS interaction energy of the different mutants was evaluated from various molecular dynamics snapshots and only variants with lower energy than the wild-type were submitted to DFT-based QM/MM calculations in order to estimate the energy barrier. The results from this virtual screening were tested and a variant with approximately 2000-fold activity improvement was obtained. In a similar way Gu *et al.* ran a virtual screening to design a mandelate racemase (EC 5.1.2.2) but a faster protocol was designed by only keeping the interaction energy evaluation step and by skipping the QM/MM energy barrier calculation.[212] In their work the binding energy of enzyme-TS complex of the mutants were evaluated by combining a molecular dynamics sampling to a molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) calculations. The best scoring mutants were experimentally tested and after two rounds of screening a positive variant with 5.2-fold improved catalytic efficiency was found. Because TS are not always available they can be approximated by transition state analogues (TSA). OptZyme[213] follows this strategy; mutation effects are evaluated by comparing the enzyme-substrate (ES) and the enzyme-TSA interaction energies. If the energy of latter is minimized rather than the former, then the mutation is predicted to be beneficial.

2.3 Summary and conclusion

In this chapter, the significance of biocatalysis for the synthesis of pharmaceuticals and the recent developments in this area have been introduced. Two main reasons motivate the use of biocatalysis in the pharmaceutical industry: the development of greener process and the easy access to chiral building blocks. As a matter of fact, biocatalysis is increasingly used for the synthesis of pharmaceuticals intermediates with important industrial relevant examples.[79, 215] Current protein engineering methods are able to successfully modify enzymes for efficient synthesis of pharmaceuticals.[216] The development of semi-rational methods have allowed more efficient designs than initially used random strategies. Nevertheless, the current time-scale is still not ideal for a perfect integration into the drug discovery and development pipeline. As stated in a recent review there is a "Need for speed".[7] The next challenge is to produce desired enzymes in reduced time-scale to align with business needs.[7] Ideally the process should be ready at product launch in order to avoid a refile with regulatory agencies. The sitagliptin example took one year and it has been suggested that doubling the speed should be sufficient.[7] The development of ultrahigh-throughput screening/selection methods in industrial environment and the integration of (new) computational methods to predict the impacts of mutations could help reach this goal. Finally, this chapter has mainly focused on chiral transformations, but robust biocatalyst platforms for other transformations relevant for the pharmaceutical industry are still in development. For example, enzymatic options for amide bond formation for industrial applications are still limited.[217]

3. Theory and Methods

Enzymes are large molecules, for example, hen egg-white lysozyme is 14.3 kDa and is considered to be a small enzyme.[218] This means that modelling the reactions that enzymes catalyse is complex and challenging. This can be complicated further by the need to include part of a particular enzyme's surrounding environment, such as the solvent, cofactors, other proteins, a lipid membrane, or DNA. The typical approach for studying systems on the scale of enzymes is through all atom molecular mechanics (MM) methods, where the electronic degrees of freedom of the molecules are ignored and only motions of the nuclei are calculated. MM is typically well parameterised for ground state properties such as geometries and interaction energies. However, this method cannot account for bond breaking and forming nor for charge transfer, and electronic excitation, that is, for any properties that are dependent on changes in the electronic structure.

Electronic rearrangement in a system can be studied by quantum mechanics as this level of theory retains the electronic structure of a system. Unfortunately, QM methods are limited to a range of tens to hundreds of atoms. From these limitations came the idea of merging the two methods (QM and MM), as suggested in the pioneering work of Warshel and Levitt.[219] Within a QM/MM framework, the substrate and enzyme atoms needed for the reaction, are modelled at a QM level, while the rest of the system is modelled at the MM level.

This chapter first introduces the basis of quantum mechanics and molecular mechanics methods. Subsequently, the quantum mechanics/molecular mechanics methods are introduced with an emphasis on the methodology used in this work.

3.1 Quantum mechanics methods

Quantum mechanics models the atom as a nucleus surrounded by electrons. The total energy of the system, can be obtained by solving the Schrödinger equation.[220] The Schrödinger equation of a multinuclear, multielectron system is expressed in **Eq. 3.1**; where $\hat{H}$ represents the Hamiltonian, Ψ is the wavefunction for the system and E the energy associated with this wavefunction. The absolute square of the wavefunction, when integrated over space and time, yields the probability of finding the particle at that position at that time.

$$\hat{\mathrm{H}}\Psi = E\Psi \qquad (3.1)$$

The Hamiltonian expression is detailed in **Eq. 3.2**, where the first two terms describe the kinetic energy of the nuclei, *A*, and the electrons, *i*, respectively, and the last three terms describe Coulombic interactions between particles. h is the Planck's constant and e is the electron charge. M and m are the nuclear and electron masses, respectively, and R_{AB}, r_{ij} and r_{Ai} are distances separating nuclei *A* and *B*, electrons *i* and *j*, and electrons *i* and nuclei *A*, respectively.

$$\hat{\mathrm{H}} = \frac{-h^2}{8\pi^2}\sum_A^{nuclei}\frac{1}{M_A}\nabla_A^2 - \frac{h^2}{8\pi^2 m}\sum_i^{electrons}\nabla_i^2 - e^2\sum_A^{nuclei}\sum_i^{electrons}\frac{Z_A}{r_{Ai}} + e^2\sum_A^{nuclei}\sum_{>B}\frac{Z_A Z_B}{R_{AB}} + e^2\sum_i^{electrons}\sum_{>j}\frac{1}{r_{ij}} \qquad (3.2)$$

The Schrödinger equation may be solved exactly for a one-electron system. However, when the system comprises more than one electron it becomes difficult to solve this equation exactly and as such there is a need for an approximate solution to the Schrödinger equation for these many-electron systems. In computational chemistry, different methods are used to determine an approximate solution to the Schrödinger equation and they can be distinguished by the approximations that they use and thus the level of accuracy that they can reach.[221]

First, *ab initio* methods make calculations from the first principles of quantum mechanics. Hartree-Fock (HF)[222] is the most basic *ab initio* method and is typically the starting point of the majority of other wavefunction based methods. HF estimates the wavefunction by applying three approximations to the Schrödinger equation. First, the Born–Oppenheimer approximation assumes that nuclei do not move and as such is an approximation to the full Hamiltonian.[223] This first approximation removes the kinetic energy of the nuclei and the Coulombic interaction between the nuclei becomes constant, as there is no variation in R_{AB}, which results in the "electronic" Schrödinger equation. The second Hartree-Fock approximation involves the approximation of the wavefunction as a set of single-electron wavefunctions, which constitute the total wavefunction. Each electron is considered independently within a field generated by the remaining other electrons. This is an important approximation because it simplifies the many-electron Schrodinger equation, which cannot be solved exactly, to a product of one-electron equations that can be solved. Finally, the linear combination of atomic orbitals (LCAO)[224] approximation, introduces

the use of predefined linear combinations of basis set to calculate molecular orbitals (**Eq. 3.3**). In Eq 3 ϕ_i are the basis functions and c_i the associated coefficients.

$$\psi = \sum_{i=1}^{N} c_i \, \phi_i \qquad (3.3)$$

The HF method is also known as a self-consistent field method (SCF). This is because the HF method requires the charge distribution to be "self-consistent" starting from an assumed initial field. Therefore, several iterations are performed in order to optimize the orbitals with the aim of minimizing the HF energy. This strategy is based on the variational principle. The variational principle expression is detailed in **Eq. 3.4**, where ψ is the normalized trial solution and E_0 is the energy of the ground state.[225] This equation states that an approximated wavefunction has an energy that is above or equal to the exact energy E_0. The method consists in choosing a "trial wavefunction" depending on one or more parameters. These parameters are adjusted until the energy of the trial wavefunction is minimized. Within the LCAO approximation these parameters are the expansion coefficients c_i in **Eq 3.3**.

$$E_0 \leq \langle \psi | H | \psi \rangle \qquad (3.4)$$

HF does not account for the correlation between the spatial positions of electrons due to their Coulomb repulsion. Thus more sophisticated *ab initio* calculations, collectively termed post-Hartree–Fock methods, will increase the level of correlation that will be associated with an increase in the level of accuracy. However, the problem with these methods is the exponentially increasing computational effort that results from including

successively more electron correlation, which makes it virtually impossible to apply these approaches efficiently to larger, more complex systems.[226]

In density functional theory (DFT) the energy is expressed as a function of the density of electrons over space. In 1964, Hohenberg and Kohn (HK) demonstrated that the ground state energy could be predicted from the electronic density.[227] The main advantage of DFT theory is that it includes correlation and scales as N^3; and thus brings in correlation at a much cheaper cost than wavefunction methods. For example Møller–Plesset perturbation theory (MP2), the first correlated HF method, scales as N^4/N^5 and CCSD(T)[228-230] as N^6. Thus using the electron density significantly speeds up the calculation. Nevertheless, in the HK formulation of the DFT, the largest source of error was in the representation of the kinetic energy. This deficiency was largely remedied by the currently use DFT introduced by Kohn-Sham in 1965 (KS-DFT);[231] by transforming the problem into a non-interacting system in an effective potential, the kinetic energy functional of this system was exactly known. This non-interacting system is made of single-particle wave functions, the density is thus expressed in terms of orbitals (**Eq. 3.5**). φ_μ are the so-called Kohn-Sham orbitals.

$$\rho(r) = \sum_{\mu=1}^{N} \left|\phi_\mu(r)\right|^2 \qquad (3.5)$$

Within KS-DFT the ground-state electronic energy, E, is written as a sum of the kinetic energy of a fictitious system of non-interacting electrons, T_{ni}, the electronuclear interaction energy, V_{ne}, the classic electron-electron Coulombic repulsion energy, J, and the exchange/correlation energy, E_{XC} (**Eq. 3.6**).

$$E[\rho(r)] = T_{ni}[\rho(r)] + V_{ne}[\rho(r)] + J[\rho(r)] + E_{XC}[\rho(r)] \qquad (3.6)$$

However, the problem with KS-DFT is that the exact form of the exchange-correlation (XC) functional is not known so that a functional that approximates this contribution to the Hamiltonian must be used. There are a multitude of different density functionals that have been developed to approximate the exact XC functional and the DFT method is therefore often referred to by the name of the XC functional that is employed.[232] The simplest approach to obtain the XC is the Local-density approximations (LDA). LDA assumes that the exchange correlation energy can locally be approximated by the exchange correlation energy of a uniform electron gas of same density. To improve the results information about the gradient of the charge density can be supplemented, this is known as the generalized gradient approximation (GGA).

While there are many pure density functionals that exist, one common approximation for the exchange component of this functional is to include a mixture of exact HF exchange with a density-based exchange functional. These types of functionals that include HF exchange with density based exchange are known as hybrid functionals. For chemical applications, the most popular hybrid functional is B3LYP,[233-234] although the Minnesota functionals (M05,[235] M06,[236] *etc.*) are becoming increasingly popular due to their accuracy across a range of applications.[68] The full B3LYP expression is given in **Eq. 3.7**, where E_X^{HF} represents the Hartree-Fock contribution (20%) to exchange, E_X^{LDA} the LDA contribution (80%) to the exchange and E_X^{B88} the gradient correction to the exchange. E_C^{VWN} and E_C^{LYP} are two correlation functionals.

$$E_{XC}^{B3LYP} = 0.2E_X^{HF} + 0.8E_X^{LDA} + 0.72E_X^{B88} + 0.19E_C^{VWN} + 0.81E_C^{LYP} \quad (3.7)$$

One well known limitation of the DFT, if uncorrected, is the inability to describe long-range correlation, which is required to account for the attractive part of the van der Waals forces, the dispersion.[237] Dispersion is an important force because it plays an important role in many chemical systems; for example, dispersion forces are important for the structures of DNA and proteins (mostly through stacking).[238] Different methods to correct this limitation exist and have been reviewed elsewhere.[239] One of the most common approaches is dispersion-corrected DFT, termed 'DFT-D'.[240] The principle of DFT-D methods is to use conventional functionals and add empirical dispersion terms (explicit attractive terms between all atomic pairs). For example, in B3LYP-D3, the usual B3LYP functional is combined to the D3 dispersion correction energy term.[241]

The effect of dispersion in enzyme-catalysed reactions has for example been studied in cytochrome P450 enzymes for which the inclusion of empirical corrections was associated with a gained in accuracy in the calculated energies while it had a small effect on the optimized geometries.[242] As the addition of such correction has very small computational cost it is recommended to use it whenever it is possible.[242]

3.2 Molecular mechanics methods

The classical potential-energy function of molecular mechanics is calculated as a sum of bonded and non-bonded terms. These interactions are modelled by a force field (an appropriate mathematical function and the associated parameters). These parameters are determined as much as possible from experimental data and supplemented with data obtained from *ab initio* calculations.

A simple, class-I MM energy expression of this type is shown in **Eq. 3.8**.

$$E_{MM} = \sum_{bonds} k_d(d - d_0)^2 + \sum_{angles} k_\theta\,(\theta - \theta_0)^2 + \sum_{dihedrals} k_\emptyset\,[1 + cos(n\emptyset + \delta)] + \sum_{nonbonded\ pairs\ AB} \left\{ \varepsilon_{AB} \left[\left(\frac{\sigma_{AB}}{r_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}} \right)^{6} \right] + \frac{1}{4\pi\varepsilon_0} \frac{q_A q_B}{r_{AB}} \right\}$$

(3.8)

The symbols d, θ, and $\emptyset$ designate the instantaneous bond lengths, angles, and torsions, respectively; d_0 and θ_0 are the corresponding equilibrium values; n and δ are the torsional multiplicity and phase, respectively. The bonded force constants are k_d, k_θ, and $k_\emptyset$; r_{AB} is the non-bonded distance between atoms A and B, and ε_{AB} and σ_{AB} are the Lennard–Jones parameters (described in detail below); q_A, q_B are atomic partial charges; and ε_0 is the vacuum permittivity (dielectric constant).

Force fields that utilise the basic potential energy expression described in **Eq. 3.8** are termed class-I force fields. Class-I force fields have been well-parameterised to represent

the majority of the biomolecular space – available examples include OPLS-AA[243], AMBER[244] and CHARMM[245].

The first three summations in **Eq. 3.8** are bonded terms. They describe the bonds, angles and dihedral rotations (torsions) of atoms that are covalently bonded in a molecule. These are illustrated in **Figure 3.1**. The bonded terms are represented by harmonic potentials for bond stretching and angle bending, and a cosinus function describe the energy profile for atoms rotating around a given bond.

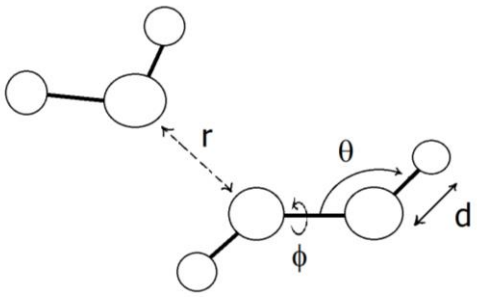


Figure 3.1. Example of parameters in classic force field.

The non-bonded terms are modelled using Lennard–Jones-type van der Waals terms and Coulomb electrostatic interaction terms, which are calculated for pairs of atoms separated by three or more bonds and between atoms in different molecules.

The van der Waals (vdW) component of the potential is:

$$E_{vdW} = \varepsilon_{AB}\left[\left(\frac{\sigma_{AB}}{r_{AB}}\right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}}\right)^{6}\right]$$

where r_{AB} is the non-bonded distance between atoms A and B, and ε_{AB} and σ_{AB} are the Lennard–Jones parameters. ε is a parameter determining the depth of the potential well and σ is a length scale parameter that determines the position of the potential minimum. The van der Waals component of the potential contains an attractive and a repulsive term (**Figure 3.2**). The physical origin of the r^{-6} attractive term lies in the dispersion forces generated between instantaneous dipoles, which arise from fluctuations in electronic charge distributions in all molecules. The quantum mechanical origin of dispersion forces was first explained by London.[246] The repulsive r^{-12} term has a quantum origin in the interaction of the electron clouds with each other (Pauli exclusion) in addition to the internuclear repulsions. The 6-12 Lennard-Jones (LJ) potential for the van der Waals interactions is a compromise between accuracy and computational efficiency.

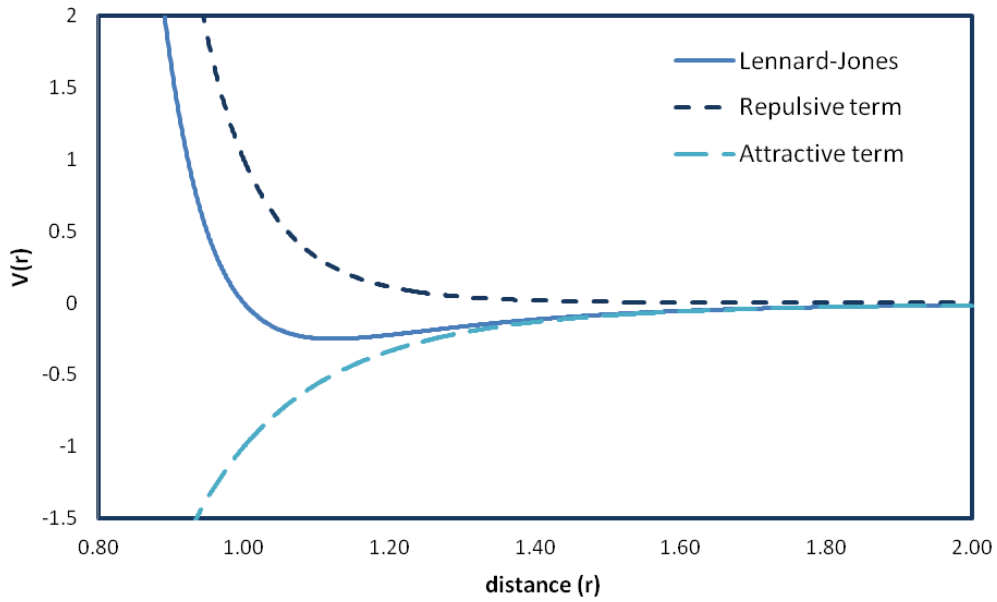


Figure 3.2. A graphic representing the different components of the Lennard-Jones potential.

The Coulomb potential is used to model the electrostatic interaction (E_{el}) between the partial charges assigned to each atom:

$$E_{el} = \frac{1}{4\pi\varepsilon_0}\frac{q_A q_B}{r_{AB}}$$

r_{AB} is the non-bonded distance between atoms A and B; q_A, q_B are atomic partial charges; and ε_0 is the vacuum permittivity (dielectric constant).

In the MM force field, electrostatic interactions are calculated between atom-centered point charges. Therefore, a lot of effort has been put into developing methods to determine partial charges that reproduce electrostatic properties of molecules, and in particular the

electrostatic potential obtained from quantum mechanical calculations.[246] Relative to the Lennard Jones potential, the Coulomb potential is long range and decays slowly.

More complex force fields include additional, higher-order terms. Class-II force field have anharmonic terms (*e.g.*, through the use of Morse potentials or quadratic terms) and explicit cross terms to account for the coupling between coordinates. The presence of these cross terms tend to improve the ability of the force field to predict properties of more unusual systems. Class-III force fields take account of chemical effects and other features such as electronegativity and hyperconjugation. In general, by adding extra terms to a potential energy expression one obtains a more accurate force field. However, parameterization of the additional parameters is more complicated, time consuming, and generally implies a lower transferability of the force field.

3.3 QM/MM methodology

Since the seminal contribution of Warshel and Levitt more than forty years ago,[219] QM/MM methods have mature to a robust methodology for biomolecular applications. More particularly, one milestone was reached in 2006 when Claeyssens and co-workers demonstrated that by using high level quantum method (*i.e.*, LCCSD(T0)) the predicted barriers for enzyme-catalysed reactions could reach near chemical accuracy (*i.e.*, within 1 kcal/mol).[72, 247] Today, QM/MM methods are a popular class of methods that find various biochemical applications such as enzymatic reactions studies, protein-ligand binding predictions or photochemistry studies.[248-251] This section introduces the QM/MM method with emphasis on the methodology used in this thesis. The software used for the QM/MM calculations in this work is Qsite.[252-253]

3.3.1 QM/MM partitioning

QM/MM methods divide the full system (*FS*) into two regions, the inner region (*IS*) that comprises the reactive site and the outer region (*OS*), which comprises the rest of the protein and the environment (*e.g.*, water, ions, *etc.*, **Figure 3.3**). The atoms that comprise the inner region are treated by QM methods, while the outer region is treated with an empirical MM potential.

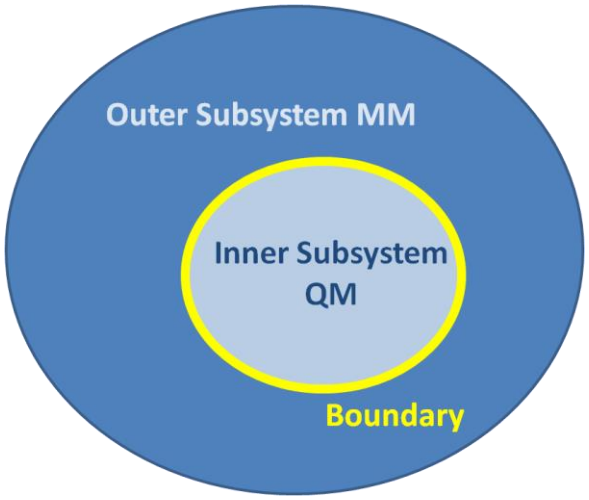


Figure 3.3. Representation of a QM/MM system.

Between the inner and outer subsystem there is a boundary region that needs to be carefully defined and needs some additional treatment when the boundary is between two covalently bonded atoms. In this thesis QM and MM atoms directly implicated in a QM/MM frontier bond are called Q1 and M1 respectively (**Figure 3.4**). QM atoms directly linked to Q1 are called Q2, those linked to Q2, are called Q3, *etc*. The same scheme is also used for MM atoms (*i.e.*,M1, M2, M3…).

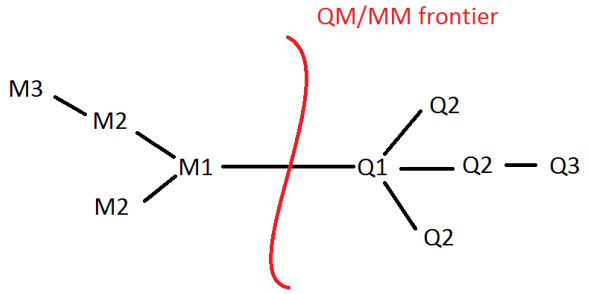


Figure 3.4. QM/MM frontier atoms annotation.

When it comes to cutting bonds during the QM/MM partitioning, there are a couple of "best practice" recommendations to limit artefacts that could be associated with this kind of partitioning.[251, 254] The QM zone should be as big as possible, although this is limited by the computational efficiency, which places a practical restriction of around 100 atoms; the boundary region should not cut through polarized bonds where possible; and finally, QM atoms participating in bond making or breaking should not be involved in any bonded coupling term, since dihedral angles are defined by three bond vectors, those QM atoms should be at least three bonds away from the boundary.[251]

3.3.2 QM/MM Energy Expressions

In the following, the notation, $E_Y(X)$ stands for the energy of X, computed at the Y level of theory. There are two main schemes for computing the total QM/MM energy: the additive and the subtractive approach. The additive scheme has been employed for the calculations presented in this thesis. This scheme adds the MM energy of the outer region, to the QM energy of the inner region and a coupling term ($E_{QM-MM}(IS, OS)$) that represent the interactions between the two systems, as shown in **Eq. 3.9**.

$$E_{QM/MM}^{add}(FS) = E_{MM}(OS) + E_{QM}(IS) + E_{QM-MM}(IS, OS) \quad (3.9)$$

The majority of QM/MM models presently in use employ the additive scheme.[254] Indeed the explicit expression of the coupling term is an interesting tool as it allows the user to control the way the inner and outer regions interact and therefore the level of accuracy of the model. Additive QM/MM models will thus differ primarily in how the coupling term is calculated and how the boundary region is treated. A correct expression of the coupling

term is needed to determine the QM/MM Energy. As shown in **Eq. 3.10**, the coupling term represents the sum of the energies from the boundary regions, the van der Waals and electrostatics interactions.

$$E_{QM-MM}(IS, OS) = E^{b}_{QM-MM} + E^{vdW}_{QM-MM} + E^{el}_{QM-MM} \quad (3.10)$$

3.3.2.1 The Electrostatic QM-MM Interaction

Electrostatic QM-MM interactions are defined as occurring between the QM charge density and the single point charge model used in MM model. Due to their long-range action, electrostatic interactions energies are typically the most significant component of the coupling term. There are different ways of calculating them depending on the level of mutual polarization of QM and MM charge model: mechanical, electrostatic or polarizable embedding.[251] In the work described in this thesis, electrostatic embedding was used for all calculations. In this approach the electrostatic effect of the environment on the QM wave function is included by adding the MM point charges as one-electron terms into the QM Hamiltonian. Karplus and co-workers[255] were among the first researchers to implement this scheme. The electrostatic embedding Hamiltonian is expressed in **Eq. 3.11**, r_i and R_j represent the position of electron i and MM atom J, h_i^{QM} is the original one-electron operator energy of electron i and M is the number of MM point charges and Q_j is the MM partial charge at the position R_j.

$$h_i^{QM-MM} = h_i^{QM} - \sum_J^M \frac{e^2 Q_j}{4\pi\varepsilon_0 |r_i - R_j|} \quad (3.11)$$

Electrostatic embedding is relatively efficient and easy to implement, which has led to its widespread use. Nevertheless, while this approach allows the MM region to polarise the QM region, it does not allow the MM region to respond to the QM region (*i.e.*, mutual polarisation).

The process of mutual polarisation is the next level of sophistication in electrostatic coupling and requires a polarisable force field. In the polarization embedding scheme both regions can mutually polarize each other, which leads to a computationally demanding approach whereby the MM and QM polarization needs to be computed in a self-consistent-field iteration. In addition to the computational demands, the largest restriction to this approach is the lack of a readily available general purpose polarisable force field for bimolecular simulation.[256]

3.3.2.2 QM-MM van der Waals Interactions

With respect to the QM-MM van der Waals coupling interactions term, the additive schemes account only for the van der Waals interactions between one atom from the inner part and one atom from the outer part. The QM-MM van der Waals is added explicitly and are typically described by a Lennard–Jones potential.[254] The only difficulty is that ε and σ parameters need to be found for the inner region atoms. These are usually taken from force field parameter for the same atoms but it may not always be a correct approximation as the character of an atom may change during the reaction.[254] Nevertheless, while every atom of the QM region is involved in van der Waals interactions

with all the atoms of the MM region, only those closest to the boundary contribute significantly.

3.3.2.3 QM/MM Boundary Treatment

For a reaction involving relatively small molecules in solution, the partitioning of the atoms in a system between different regions is simple – the molecules implicated in the reaction are put in the QM region and the solvent molecules constitute the MM region. The situation is more complicated when the region of interest is part of a large molecule, *e.g.*, the active site of an enzyme. In that case, parts of the same molecule may be in different regions and there may be covalent bonds between the QM and MM atoms. It is not possible to simply truncate these bonds as this would leave half-filled orbitals and give an inaccurate description of the electronic state of the QM region. Therefore, an appropriate technique to treat these "dangling bonds" is required.[100]

The link atom approach has been used in this thesis. It was introduced by Singh and Kollman.[257] Within this approach a "link atom" is employed to cap the QM region to avoid leaving dangling bonds. This is done by the introduction of a monovalent link atom (LA, typically hydrogen) along each QM/MM bond (**Figure 3.5**). The LA replaces the MM atoms of the broken bonds in the QM calculation. The link atom is only present in the QM calculations and is not seen by the MM atoms. The bond Q1–M1 is described at the MM level. QM calculations are then performed on an electronically saturated system consisting of the inner subsystem and the link atom(s), *IS*+LA.

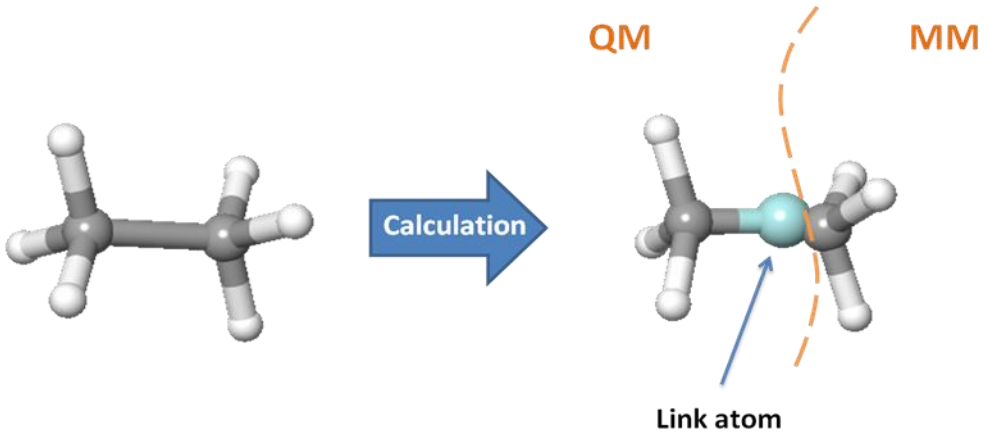


Figure 3.5. Link atom method.

Apart from the fact that the link atom is chemically and electronically different from the group that they replace, they also create additional degrees of freedom in conformation and molecular orbitals. To remove the additional conformational degrees of freedom, most of the current link-atom schemes scale the position of the link atom to the positions of Q1 and M1, such that that the additional 3 degrees of freedom are removed.[258-260] LA is therefore transparent during the geometry optimization as this process only handles independent variables. A link atom also introduces a risk of over-polarization as the LA results in an extension of the QM density closer to the boundary MM atoms (*i.e.*, M1).[260] To avoid over-polarization, a solution comprising deletion or redistributing of the point charges in the boundary region is typically employed.[260] Within the QSite QM/MM approach, the redistribution of charges at the boundary region is handled by using Gaussian charge distributions represented on a grid to represent the potential of the M1 atoms and MM point charges are employed for the rest of the MM region. Another way to treat boundary bonds is the use of hybrid orbitals. Currently, there are two methods of

this kind: localized SCF method (LSCF)[261] and the generalized hybrid orbital approach (GHO).[262]

3.3.3 Optimization techniques for QM/MM systems

Optimization of large systems is not a straightforward task due to the large number of structural degrees of freedom. For instance the full MM optimization of a protein will typically require more than 1,000 steps.[263] This is a big issue for QM/MM optimizations as routinely running 1,000 QM SCF cycles is prohibitive. Thus, particular strategies must be adopted to make optimization affordable. One strategy can be to take advantage of the QM/MM partitioning and run the optimization of the separate regions sequentially, rather than simultaneously. This approach, termed a microiterative scheme, divides the system into a core containing (at the minimum) the QM atoms and an environment, containing the rest of the system.[264] After each optimisation step of the core (the 'macroiterative' cycle), the environment is fully optimised (the 'microiterative' cycle). By optimising in this way, the number of QM evaluations is reduced significantly at the cost of increasing the number of MM evaluations of the environment. As MM evaluations are usually much cheaper, this reduces the overall computational time.

Another, complementary approach, to improve QM/MM optimization is to employ different type of coordinates and optimization algorithms in the two regions. Thus, while the minimisation of the core is done at a high level (in this work with quasi-Newton) using internal coordinates, the environment will be minimised by techniques that avoid a full

Hessian manipulation (in this work with the truncated Newton method[265]) and uses Cartesian coordinates.

Finally, when microiterative optimizations are associated with electrostatic embedding schemes, the difficulty that arises is that the electrostatic QM-MM interaction is evaluated at the QM level. Therefore, a QM calculation should be done at each MM step to let the density adapt to the new MM configuration and to obtain the forces on the MM due to the QM density. However, the relaxation during the MM optimization is neglected and the QM charge distributions are approximated by point charges.[251]

3.3.4 Model preparation

The starting point for any QM/MM calculation is to find and prepare a starting structure for the simulations. This starting structure usually comes from a X-ray crystal structure but examples of QM/MM calculations that start with homology models and giving accurate results exist.[266] To prepare the system for a QM/MM simulation several steps are needed (**Figure 3.6**). It is necessary to ensure that if there are any missing residues in the crystal structure that these are replaced or, if they are terminal residues, they can be capped. Usually hydrogens are not resolved in X-ray structures, because of their low electron density, so they should be added and their orientation optimized. Titratables residues (*e.g.*, charged residues) can adopt alternative ionization states so the right protonation state should be assigned. Some residues can adopt alternative or "flipped" conformation (*e.g.*, Asn, Gln and His), the best conformation should also be assigned. In

this work all this preparations steps have been done by the "Structure preparation"[267-268] module of the MOE[268] (Molecular Operating Environment) software.

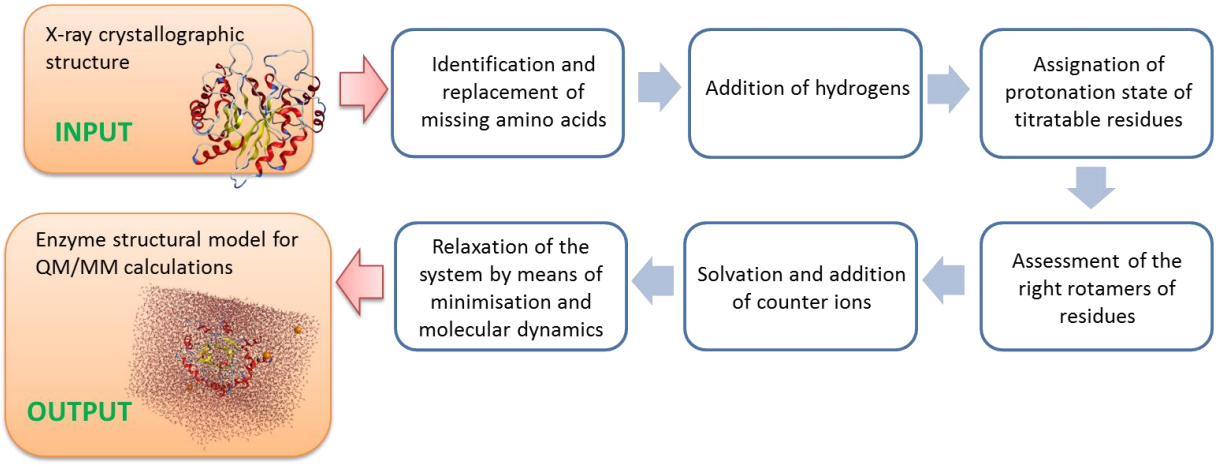


Figure 3.6. Classical procedure for X-ray crystallographic structure preparation for QM/MM calculations.

The crystal structure chosen will preferably have the substrate bound or a transition state analogue.[269] If the crystal structure contains another ligand it will need to be transformed to the desired substrate. Finally if no ligand is present the substrate will need to be docked into the active site.

As water constitutes the environment in which proteins interact, protein crystallographic structures needs to be solvated, usually in explicit water. In this work a orthorhombic box of water with a 10Å buffer between the solute and the box boundary in each direction was set up using the Desmond[270-271] integrated panel for solvation. To solvate a protein in a box of water under periodic conditions, the solvent is placed by replicating "boxes" of solvent molecules and deleting molecules whose center of mass lies outside the periodic

box boundary, and molecules that are inside or have significant overlap with the solute. Several iterations of relaxation and rehydration are carried out to obtain a more equilibrated solvent structure. The TIP4P[272] water was chosen as the water model in all work described in this thesis because it was found to be the most compatible of MM water models with all QM methods, in terms of both structural and energetic considerations, showing particularly good compatibility with the BLYP method.[273]

Following the addition of water, the system is relaxed in order to relax the added hydrogens and water molecules through several cycles of molecular dynamics; this is done gradually by first running simulations only on the water and protein hydrogens with the rest of the protein kept fixed. In this work we used the default relaxation protocol implemented in Desmond;[270-271] within 6 steps a series of short (12 and 24 ps) molecular dynamics simulations are performed to relax the model system.

A short molecular dynamics (MD) simulation of the complete system, unconstrained, can then be run to relax further the system but also to generate different snapshots that can be used as starting points for the energy barriers calculations; this ensures that the protein and solvent conformational changes are taken into account.[269] Snapshots for QM/MM calculations where selected on the basis of geometrical parameters, more specifically by measuring distances between reacting atoms.

In the work described in this thesis all the MD simulations were ran using the NPT ensemble if not defined differently in the text. The pressure is maintained at 1.01325 bar by using the Martyna-Tobias-Klein (MTK)[274] method coupled to the Nose−Hoover

thermostat.[275] The temperature are kept fixed at 300 K. A cut-off radius for the non-bonded interactions (Coulombic and van der Waals) are fixed to 9.0 Å. A tapering to zero is done for the van der Waals interactions at the cut-off. The long range Coulombic interactions in the simulation are treated by a smooth particle mesh Ewald (PME) method with a tolerance value of 1 nm.[276] At the end of the MD simulation the average value, the standard deviation, and the slope of different properties (potential energy, pressure, temperature and volume) are calculated and analyzed to confirm that the MD simulation was at equilibrium.

3.3.5 QM/MM Hamiltonian

In this thesis an hybrid QM/MM Hamiltonian was employed using Qsite.[252-254] For the classical region (MM), the OPLS2005[243] force field was always used to describe the protein. The QM region was typically modeled at the B3[277]LYP[234, 278]/6-31G*[279] level of theory, variations to this level theory are described explicitly in the text. Dispersion-corrected B3LYP are not yet available in Qsite and could not be used for this work.[280]

3.3.6 Exploration of the potential energy surface

3.3.6.1 The potential energy surface

The potential energy surface (PES) refers to the relationship between the energy of a molecule and its geometry. This geometry can be characterized by different geometric parameters (*e.g.*, bond lengths, angles, differences in bond lengths, *etc.*) and one of these parameters is the reaction coordinate. The representation of the energy along the reaction

coordinates is called the energy profile. As discussed in Section 2.1.2, the underlying effect of the enzyme is to change the free energy profile of the reaction being catalysed. Obtaining the appropriate energy profile is thus the first requirement to understanding the reaction process.

QM/MM methods can obtain the energy profile of a reaction by accurately describing the chemical changes through the redistribution of electrons within the active site. From this energy profile it is possible to identify structural and energetic information concerning the reactant, the TS, the product and any intermediates along the reaction path. Each of these structures is a stationary point within the PES. More precisely, TSs are a first-order saddle point (maximum along one direction – the reaction coordinate) while the other species are minima with respect to all coordinates.

3.3.6.2 PES exploration QM/MM methodology

In this work the PES was explored via QM/MM geometry optimisation.[281] Calculations start with equilibrated and QM/MM optimized ES complex. Alternatively, if a crystal structure with a transition state analogue bounded to the active site is available it can also be used as starting point to the study of the catalytic mechanism of the enzyme. TS search calculations are then done in Qsite[252-253] using the *Standard* method. This takes an initial guess of the TS as input and tries to find the closest saddle point to it by maximizing the energy along the lowest-frequency mode of the Hessian and minimizing the energy along all other modes. Initial guesses are built by small modifications of the optimized structure of the reactant, the goal being to make it resemble as much as possible the believed

transition state. The nature of structures can be confirmed by quantum chemical vibrational analyses;[282-283] a set of vibrational normal modes and frequencies are obtained by first calculating and then diagonalizing the full Hessian matrix. The Hessian matrix contains the second derivatives of the potential energy with respect to the coordinates. TS structures have a single imaginary frequency because they are true saddle points on the PES. Reactant and products have only positives frequencies.

After identification of the TS, this is minimised to verify that is truly connected to the reactant and to identify the next saddle point in the path (either an intermediate or the product). This step can usually be done by an intrinsic reaction coordinate (IRC) calculation. However, although IRC is implemented in Qsite[252-253] it is not operational and could not be used for this work.

This procedure is repeated for all transitions steps until the full reaction path is identified.

3.3.6.3 Analysis of the results

Once all the stationary points along the energy profile have been characterised it is possible to determine the relative energies (E_R) between them. Importantly, this results in the determination of the activation energy (E_a) (the relative energy between the reactant and the highest energy transition state) of the reaction. The activation energy is an important parameter because it can be used both as a qualitative and a quantitative tool to understand reaction mechanisms.

First as a qualitative tool the activation energy allows the identification of the best answer between different hypothetical mechanisms of the same reaction. Indeed, the one with the lowest activation energy will be the most probable mechanism. In the same way it can allow the *in silico* evaluation of the impact of a mutation – highlighting mutations that are predicted to decrease the activation energy could potentially increase the rate constant and thus the activity of an enzyme.

Second, the activation energy can be used as a quantitative parameter. In principle, the barrier obtained from computation (E_a) and experiment activation free energy ($\Delta G^{\ddagger}$), calculated from the rate constant of the reaction through transition state theory (**Eq. 2.2**), should be the same.[284] However, this is not always the case because the accuracy of the prediction will strongly depend on the level of QM theory used for the QM/MM calculation. An activation energy is near chemical accuracy when the error compared to experiment results (obtained from kinetic studies) is less than 1 kcal/mol and there are reported QM/MM studies that satisfy this criteria.[72] The activation energy can thus be a tool to access the quality of a QM/MM calculation.

3.3.7 Charge modification procedures

To evaluate the electrostatic impact of each residue of the protein on the reaction barrier a sequential procedure, henceforth referred to as 'ElectroScan', was applied on the selected reaction step (**Figure 3.7**). The partial charges of the side chain of a single residue were set to zero in both enzyme-reactant complex and transition state of the step. Then the relative energies were re-evaluated, based on the modified partial charges by single point

calculations so that the activation energy can be re-calculated. Similar charge nullification procedures have been applied in the past and have proven to be useful to identify important residues for enzymatic reactions.[46, 210, 285] Two other approaches, that we will call 'PosiScan' and 'NegaScan', were also tested where the partial charges of the side chain atoms of a single residue were set to zero except for the Cβ partial charge, which was set to a value of +1 or -1, respectively. Both 'PosiScan' and 'NegaScan' protocols were investigated as tool to introduce new favorable electrostatic interactions.

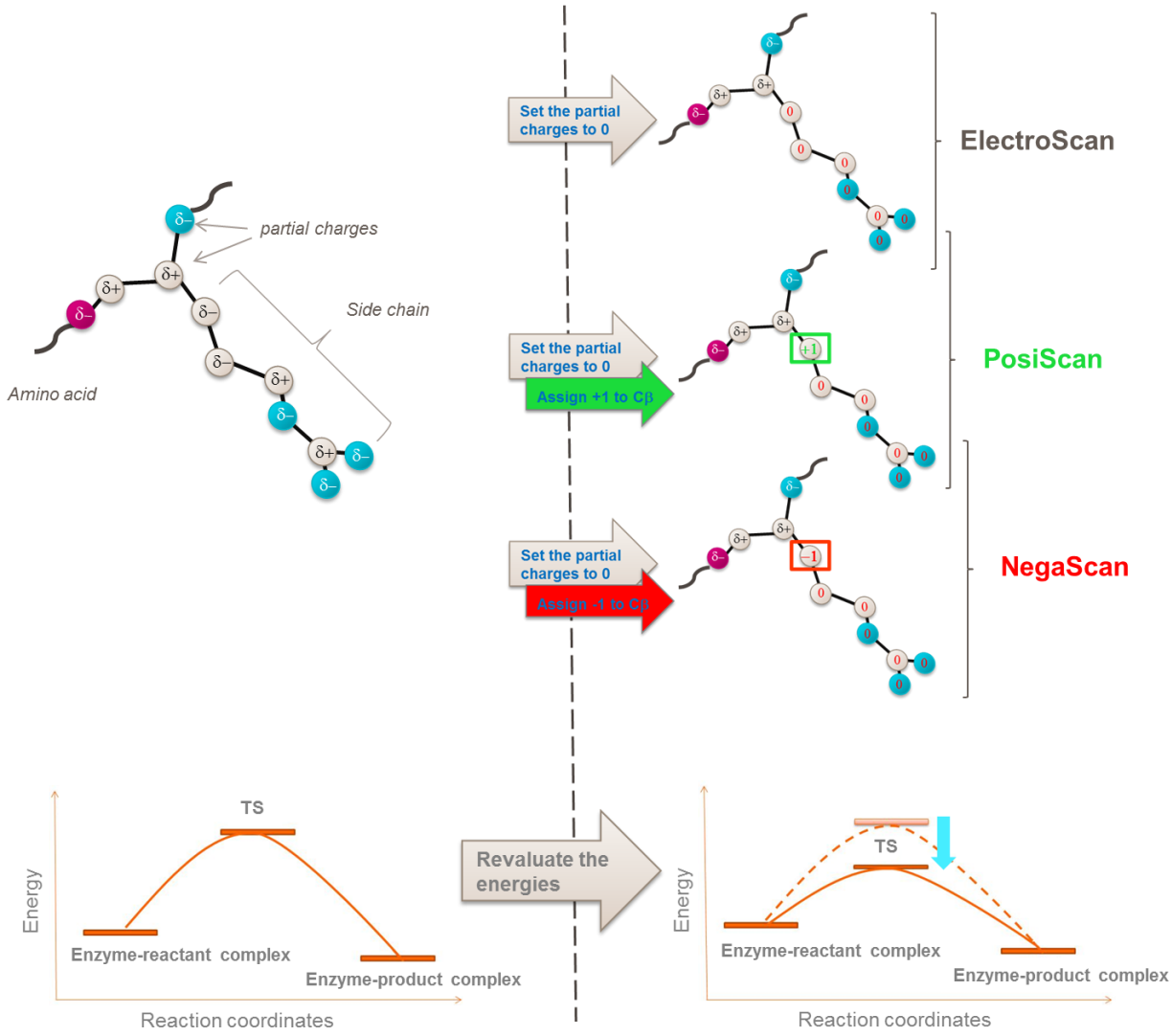


Figure 3.7. Schematic explanation of the different charge modification procedures: 'ElectroScan', 'PosiScan' and 'NegaScan'.

4. Computational investigation of Aldose Reductase catalytic mechanism

4.1 Introduction

Aldose reductase (AR, EC 1.1.1.21) is a cytosolic NADPH-dependent oxidoreductase enzyme that belongs to the superfamily of aldo-keto reductases (AKR).[286-287] Aldose reductase has been identified as the first enzyme involved in the polyol pathway of glucose metabolism which converts glucose to fructose via sorbitol.[288] This is of particular interest for the pharmaceutical industry as glucose over-utilization through the polyol pathway has been linked to tissue-based pathologies associated with diabetes mellitus complications.[288-289] AR has thus been widely studied in order to develop potent AR inhibitors to prevent or delay the onset and progression of these complications.[290] As a result, the Protein Data Bank (PDB) accounts to date (April, 2018) for an impressive number of X-ray crystallographic structures (136) of human aldose reductase.[28]

The human AR enzyme comprises 315 amino acid residues and has a β/α barrel structure (**Figure 4.1a-b**).[291] The barrel is composed of eight parallel β-strands and eight adjacent peripheral α-helical segments that are running anti-parallel to the β-sheet. The catalytic active site is located in the barrel core. The nicotinamide-adenine-dinucleotide phosphate (NADP) cofactor is situated at the top of the C-terminal end of the β/α barrel, with the nicotinamide ring projecting into the centre of the barrel and the pyrophosphate part on the border of the barrel.

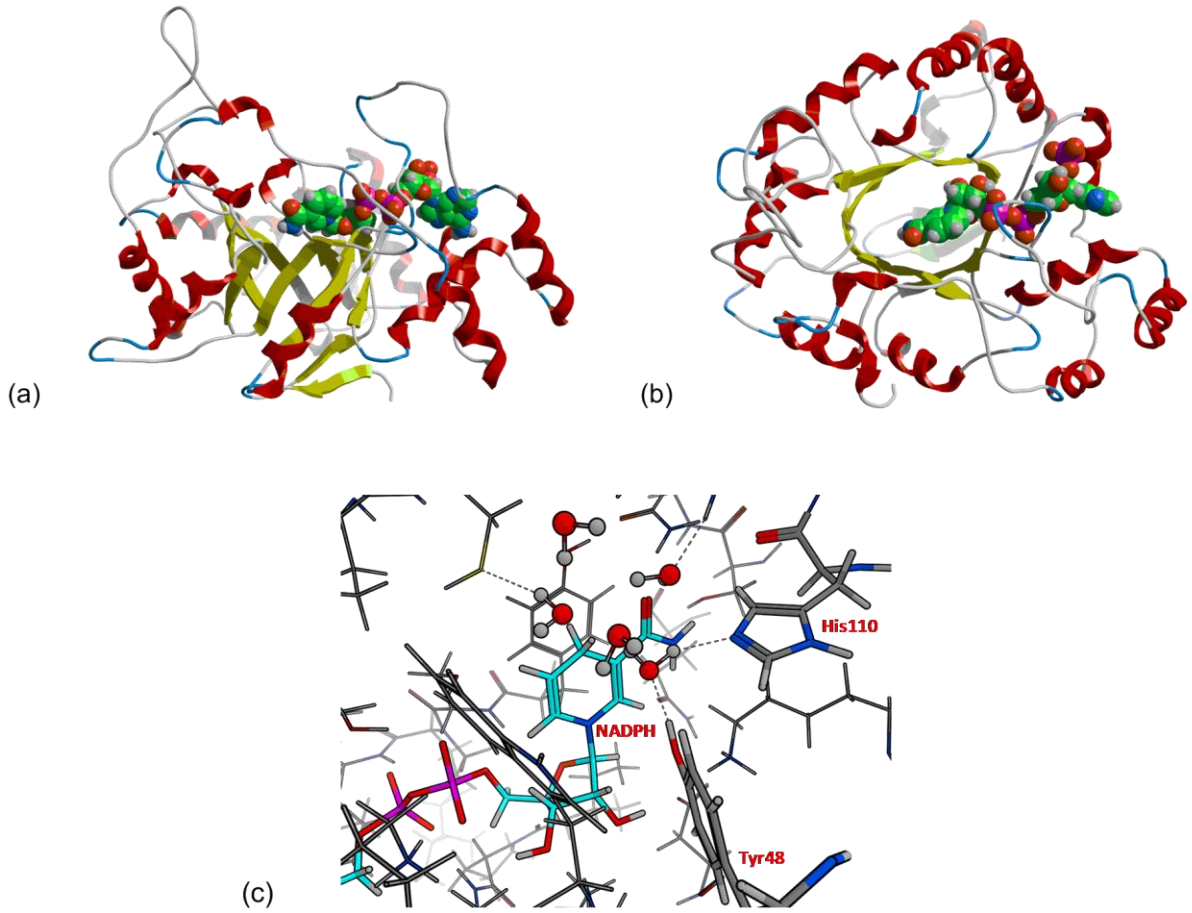
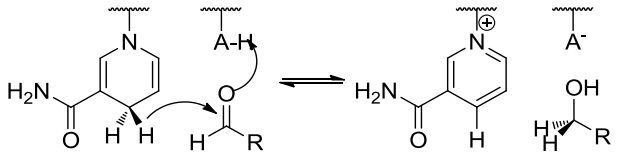


Figure 4.1. (a-b) View of the α-carbon backbone trace (schematic diagram) of the aldose reductase structure with bound NADPH. (a) View perpendicular to the β/α barrel with NADPH shown in green space-filling model (b) The structure viewed down the C-terminal end of the β/α barrel. (c) Aldose reductase active site (PDB ID: 1ADS) with crystallographic waters shown.

The reaction mechanism of aldose reductase in the direction of aldehyde reduction comprises two steps.[292] The first step is the transfer of the pro-(*R*) hydride of NADPH to the *re* face of the substrate's carbonyl carbon. The second step is donation of a proton to reduce the carbonyl to an alcohol (**Scheme 4.1**).



Scheme 4.1. Schematic representation of aldehyde reduction by aldose reductase.

Despite the acceptance of this general mechanism, several key features remain unclear. On the one hand, it is not known whether the reaction occurs in a concerted or step-wise manner. On the other hand, it is not clear which of the proximal residues, Tyr48 or His110, acts as the proton donor. Indeed, both of these residues could potentially occupy this function, as crystal structures indicate that they are well positioned to be potential proton donors during catalysis; in crystal structure 1ADS a water molecule in close proximity to the nicotinamide is hydrogen bonding to both Tyr48 and His110 and thus indicates a possible position for the substrate (**Figure 4.1c**). A comparison of the relative pK_as of the residues suggests that the lower value of histidine (pK_a=6) relative to tyrosine (pK_a=11), would make it a more likely candidate to donate a proton.[293-294] However, the proximity of the Lys77-Asp43 pair in the binding site, has been proposed to lower the pK_a of Tyr48 to 8.25 through hydrogen bonding.[294]

In the literature, there are several computational studies that investigate which of the two potential residues is the proton donor.[295-297] These include two quantum mechanics/molecular mechanics (QM/MM) studies, one by Lee and co-workers[295] and a second by Várnai and co-workers[296] and one empirical valence bond (EVB) study by Várnai and Warshel.[297] From the results of the two QM/MM studies, which are

summarized in **Table 4.1**, it is evident that they differ in both their proposed mechanisms and calculated energetics. The Lee and co-workers[295] results show a concerted mechanism while Várnai and co-workers report a step-wise mechanism.[296] The difference in the calculated energetics of the reactions is also significant, with a difference for the calculated relative energies of around 10 kcal/mol. The experimental activation free energy, determined from reaction rate studies, is 14.8 kcal/mol.[298] Thus, both computational studies overestimate the activation energy with a relative energy of 21.3 kcal/mol for Lee and co-workers[295] and 31.8 kcal for Várnai and co-workers.[296] Nevertheless, both studies agree that the reaction mechanism is more favorable with the His110 model than with the Tyr48 model, as the relative activation energies in both studies are smaller when employing His110 as the proton donor.

Table 4.1. Summary of previous QM/MM results for the reduction of D-glyceraldehyde by aldose reductase.[a,,b, c]

Study	**Level of theory**	**TS1**	**I**	**TS2**	**P**
		proton donor His110			
Lee *et al.*[295]	HF/4-31G	21.2	/	/	-12.4
Várnai *et al.*[296]	AM1	31.8	25.4	35.4	-5.9
		proton donor Tyr48			
Lee *et al.*[295]	HF/4-31G	24.3	/	/	-3.7
Várnai *et al.*[296]	AM1	41.2	33.6	34.7	10.3

[a] TS: transition state, I: intermediate and P: product

[b] Electronic energies (ΔE) in kcal/mol are given relative to the reactant state for each system studied.

[c] In Lee *et al.* study the electronic energies were obtained using a QM region with a total of 54 atoms as represented in Figure 4.2a. In the Várnai *et al.* study a larger QM region than in the Lee *et al.* study was used (the same residues are included in the QM region but the link atom positioning is different).

In the EVB study from Várnai and Warshel,[297] the energy profile was only evaluated for the tyrosine proton donor hypothesis as their detailed pK_a studies on both Tyr48 and His110 suggested that the Tyr48 proton donor hypothesis would be the most probable mechanism. More specifically, the free energy of protonation of His110 in the protein environment was evaluated and a high value of 9 kcal/mol was obtained which makes this residue very difficult to protonate in a first place. In their work the activation free energy was calculated to be 17 kcal/mol and thus in good agreement with experimental results. The better agreement of the EVB results is not a surprise as the method comprises significant sampling and is thus able to evaluate free energies that can be directly compared to experiment results. On the contrary in the two QM/MM methods described above no sampling is included and thus only potential energies are calculated. Entropic contributions are thus not included in these original calculations. While entropic and thermal contributions can play a significant role in determining transition state energies, previous work has shown that the entropic contributions to the activation energies for some enzyme reactions can be minimal,[72] and as such the underlying difference in the quality of the results between the EVB and QM/MM calculations is not necessarily due entirely to the exclusion of entropic effects.

Overall, the opposing nature of the conclusions from these two QM/MM studies, combined with the low accuracy of the calculated activation energies, indicate that a more detailed study into this important mechanism is warranted. In the present work we have examined the catalytic mechanism of aldose reductase with a QM/MM approach employing density functional theory (DFT) as the QM methodology. The structures of transition states and intermediates involved in the reaction, the energy profiles and the roles of keys residues are presented herein. The detailed interpretation of the catalytic mechanism that results from this work is helpful for the design of mechanism based inhibitors like transition-state analogue or covalent inhibitors.[299] Finally, one of the main objectives of this work is to determine how the methodological choices in a QM/MM calculation can have significant effects on both the calculated energetics and the resulting interpretation of the preferred mechanism. Therefore, the extent to which using a modern density functional and a larger QM region can affect previous results, both quantitatively and qualitatively, is also discussed.

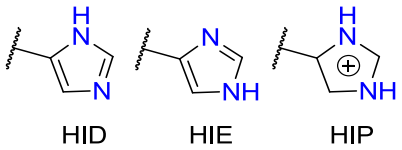
4.2 Computational methods

4.2.1 Model systems

Theoretical studies were performed starting with an X-ray crystal structure (PDB ID: 1ADS) of the aldose reductase enzyme that has a resolution of 1.65Å.[291] Although there are many other available crystal structures, this provided a convenient point of reference, as the structure was used in the previous QM/MM studies from Lee and co-workers.[295] The structure includes the cofactor NADP, so this was transformed into the reacting form NADPH. The "Structure preparation" module of MOE[268] was used to prepare the structure as explained in Section 3.3.4.

Particular attention was paid to His110, as at physiological pH, histidine can exhibit three different protonation states, that is to say HIP (protonated Nε and Nδ), HIE (protonated Nε) and HID (protonated Nδ) (**Scheme 4.2**).[300-301] Within a protein, standard pK_a values of residues can be more or less influenced by the environment and that makes the prediction of the residues' protonation state less straightforward. Different methods exist to predict the pK_a of residues but results from these predictions are not always reliable.[302] In this case we have not attempted to do a QM/MM pK_a prediction, rather the initial calculation of the protonation states was carried out with the empirical modeling program PROPKA[303]. However, the calculated protonation states from this program were found to be unreliable for the system under study with at the same time His110 predicted to be unprotonated (with a very low pK_a) and Tyr48 a very bad proton donor (because of a high

pK_a). As such, all possible protonation states for the histidine residue involved in the mechanism were evaluated.



Scheme 4.2. Different Protonation States of Histidine.

D-glyceraldehyde (GLD) was chosen as the ligand, to be consistent with the reference studies.[295-296] The accuracy of the MM force field parameters associated to GLD were tested by performing various conformational searches and minimizations with MacroModel.[304] The consistency of the bonds and angles of the resulting structures were checked using Mogul.[305] As the crystal structure did not contain any ligand, GLD was added manually. To ensure that the *re* face of the carbonyl of GLD would be able to receive the hydride from the NADPH, the carbonyl oxygen of GLD was positioned within the range of hydrogen bonding interactions with the Nε hydrogen of His110 and the hydroxyl of Tyr48. In order to get an adequate pose, an optimization with constraints on the distances of these hydrogens bonds was run to reproduce the distances from the Michaelis complex (MC) described by Lee and co-workers.[295] After having deleted all crystal waters the system was solvated as described in Section 3.3.4 of the method chapter. Finally, to neutralize the system, 3 sodium atoms were added randomly for the HIP model and 4 for both HIE and HID models.

The three model systems were gradually relaxed using a standard protocol implemented in Desmond.[270-271] A molecular dynamics (MD) simulation was performed for 1 ns using Desmond[270-271] in order to relax further the system and to obtain a variety of snapshots for the QM/MM calculations. The quality of these simulations was checked by plotting the different RMSD which results can be found in **Appendix Figure A.1.1**.

To obtain an averaged energy barrier a minimum of two snapshots per model, with suitable hydrogen bonding between GLD and both His110 and Tyr48 were extracted from the MD output and prepared for QM/MM calculations. The selected snapshots were then MM minimized to return the system to 0 K by using the truncated Newton method[265] implemented in Impact[306]. A second MM minimization using the Polak-Ribiere Conjugate Gradient[307] (PRCG) method implemented in MacroModel[304] was used in order to reproduce the distances from the Michaelis complex described by Lee and co-workers.[295] All the minimizations were done with waters' oxygen atoms kept constrained. These post-equilibration, minimized structures, which represent the enzyme-substrate (ES) complex, were used as starting structures for the QM/MM calculations.

4.2.2 QM/MM Methodology

In previous QM/MM studies from Lee and co-workers[295] and Várnai and co-workers,[296] the choice of the QM region for both studies include all hypothetical reacting species (D-glyceraldehyde, NADPH, His110 and Tyr48) and influential residues (Asp43 and Lys77).[308] An increase in accuracy can be expected if the size of the QM region is extended;[309-310] currently QM/MM calculations can readily account for up to 100 atoms

in the QM region,[311] so performing calculations on the upper side of this range could thus be considered. The QM treatment was done at an *ab initio* level (HF/4-31G) by Lee and co-workers[295] and at a semi-empirical level (AM1)[312] by Várnai and co-workers.[296] While issues such as boundary effects, the classical potential and optimization strategies may all affect calculated results, in this comparison the difference between the QM treatment of the system could be the main reason of lack of accuracy in the previous results. As such, this hypothesis will be tested in the current work. Ideally, one would perform all QM calculations with the most accurate *ab initio* method, together with the largest available basis set.[269] The most generally reliable and routinely used QM treatment in current QM/MM studies is DFT, particularly with the B3LYP functional.[277, 311] In this work the QM region was modeled at the B3[277]LYP[234, 278]/6-31G*[279] level of theory.

The effect of the size of the QM region was examined using two different QM regions on the HIP model. The first QM region was defined as in the study by Lee and co-workers.[295] This QM region with a total of 54 atoms is represented in **Figure 4.2a**. A larger QM/MM partitioning, as defined in **Figure 4.2b**, was also used in the three models (HIP, HIE and HID). In this second partitioning the same residues are included, but the QM/MM frontier is positioned differently. First, the entire side chains of residues were included by cutting between Cα and Cβ. Second, the frontier within NADPH was extended by adding the ribose part of NADPH, allowing a cut between two carbons rather than between a carbon and a nitrogen atom, consistent with best practice for the positioning of link atoms.[260] The QM/MM partitioning 2 thus had a total of 90 atoms included for the HIP model and 89 atoms for both HIE and HID models.

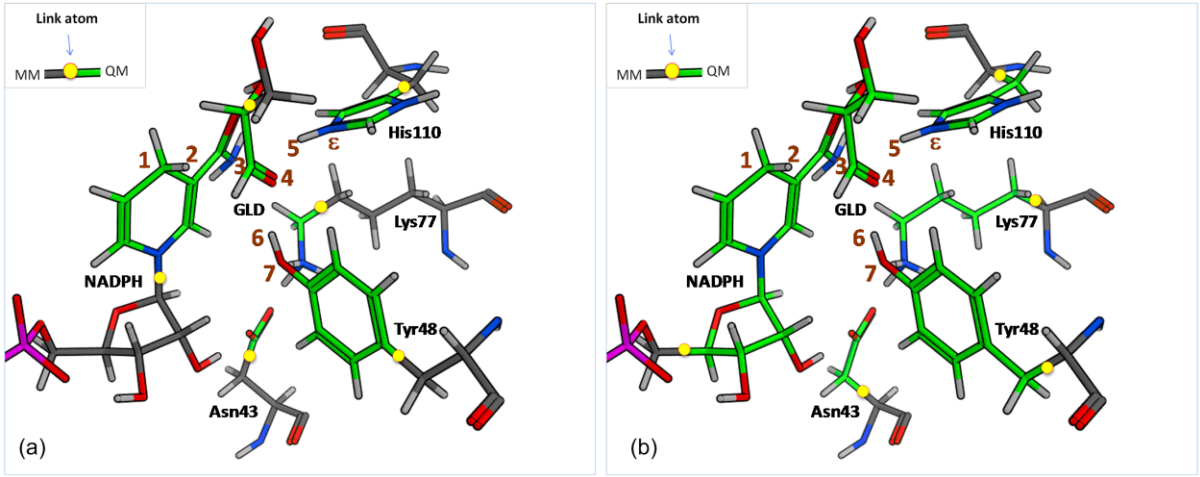


Figure 4.2. QM/MM partitioning (a) 1 and (b) 2. The QM region is defined in green. Yellow dots represent link atom positions.

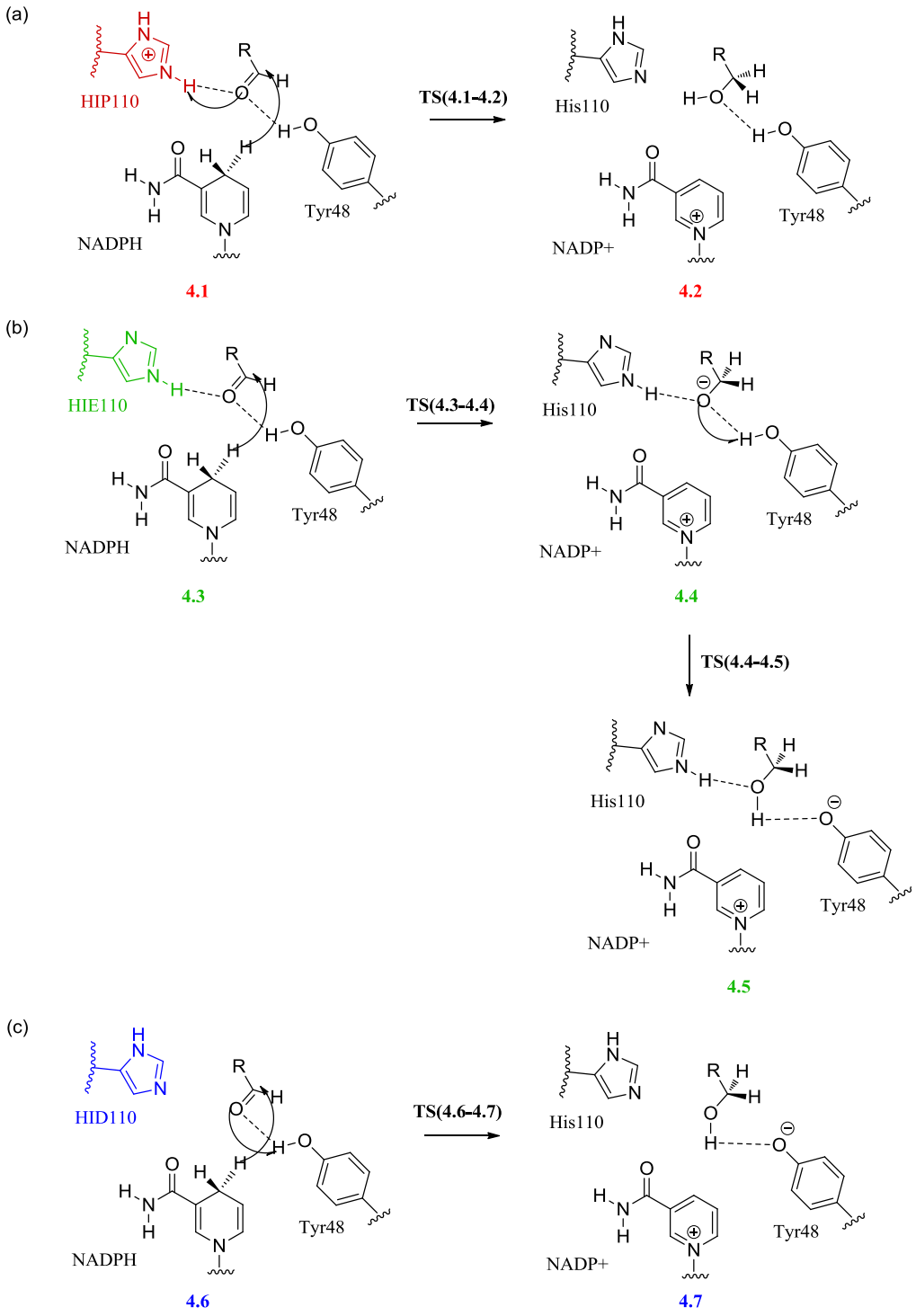
To be consistent with the reference study, the link atom approach was used to saturate the valence of the QM/MM frontiers.[257, 295]

All atoms beyond 10 Å from the reactant were consistently kept constrained during the QM/MM simulations in order to speed up the calculations. The equilibrated ES complex was optimized with QM/MM calculations. The potential energy surface (PES) for the reaction was explored starting from this optimized structure of the reactant using the approach described in the methods chapter Section 3.3.6. The initial guess for the TS was built by small modifications of the optimized structure of the reactant, the goal being to make it resemble as much as possible the believed transition state. To do so the reacting bond C1-H2 (the carbon/hydride bond of the NADPH) was elongated manually in order to position the hydride half-way between the C1 carbon where the hydride is initially attached and the GLD carbonyl carbon C3. The carboxyl double bond of the GLD was

also elongated to mimic the transition from a carbonyl double bond to an alcohol single bond. Also, in order to help the TS search process, Qsite allows one to indicate as an input what bonds are supposed to be made or broken. This was done by adding a *connect* section to the input file were C1-H2 hydride bond was defined as the reaction coordinate. In order to find the reactant and product associated with this saddle point, the TS was minimized at the same level of theory. The nature of the structures was confirmed from analysis of the Hessian.

4.3 Proposed reaction mechanism

The proposed mechanism for the different protonation states are represented in **Scheme 4.3** and the associated relative energies in **Figure 4.3**. In the following sections a detailed description of the three different reaction mechanisms is given.



Scheme 4.3. Proposed mechanisms for the different protonated states of His110 (a) HIP, (b) HIE and (c) HID.

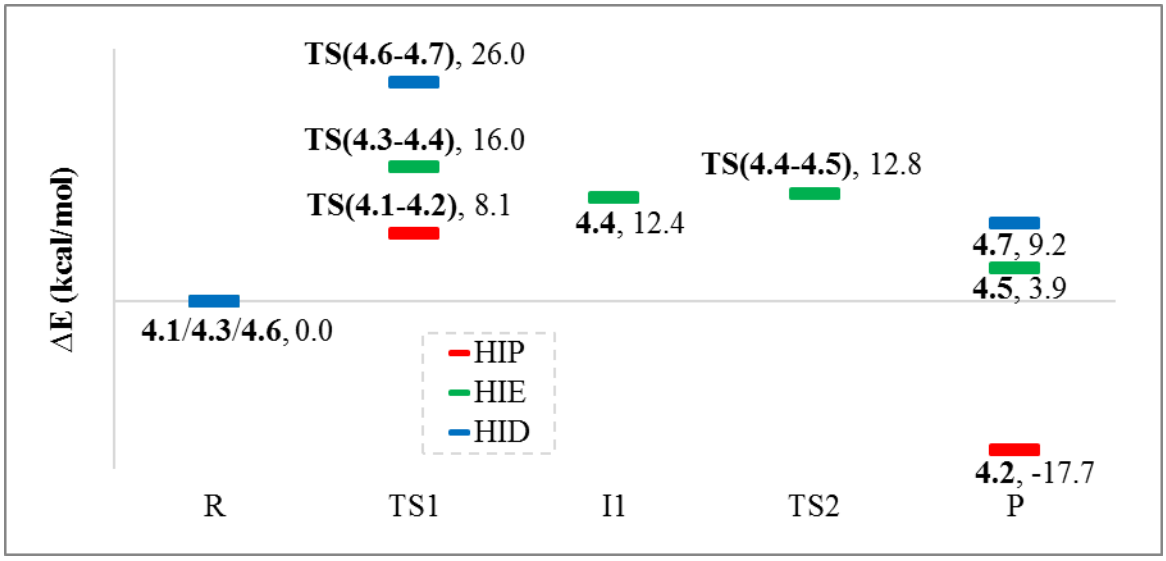


Figure 4.3. Comparison of relative energies for the three protonation model: HIP (red), HIE (green) and HID (blue). Electronic energies (ΔE) in kcal/mol, calculated at the B3LYP/6-31G* level, are given relative to the reactant state for each system studied.

4.3.1 Mechanism with HIP110

The results for the mechanism of GLD reduction by AR in the case of a protonated histidine show a single step mechanism with associated activation energy of 8.1 kcal/mol (**Scheme 4.3a** and **Figure 4.3**). A schematic representation of the starting enzyme-substrate complex **4.1**, the transition state **TS(4.1-4.2)** and the final enzyme-product complex **4.2**, including only the closest atoms around substrate, is given in **Figure 4.4a-c**.

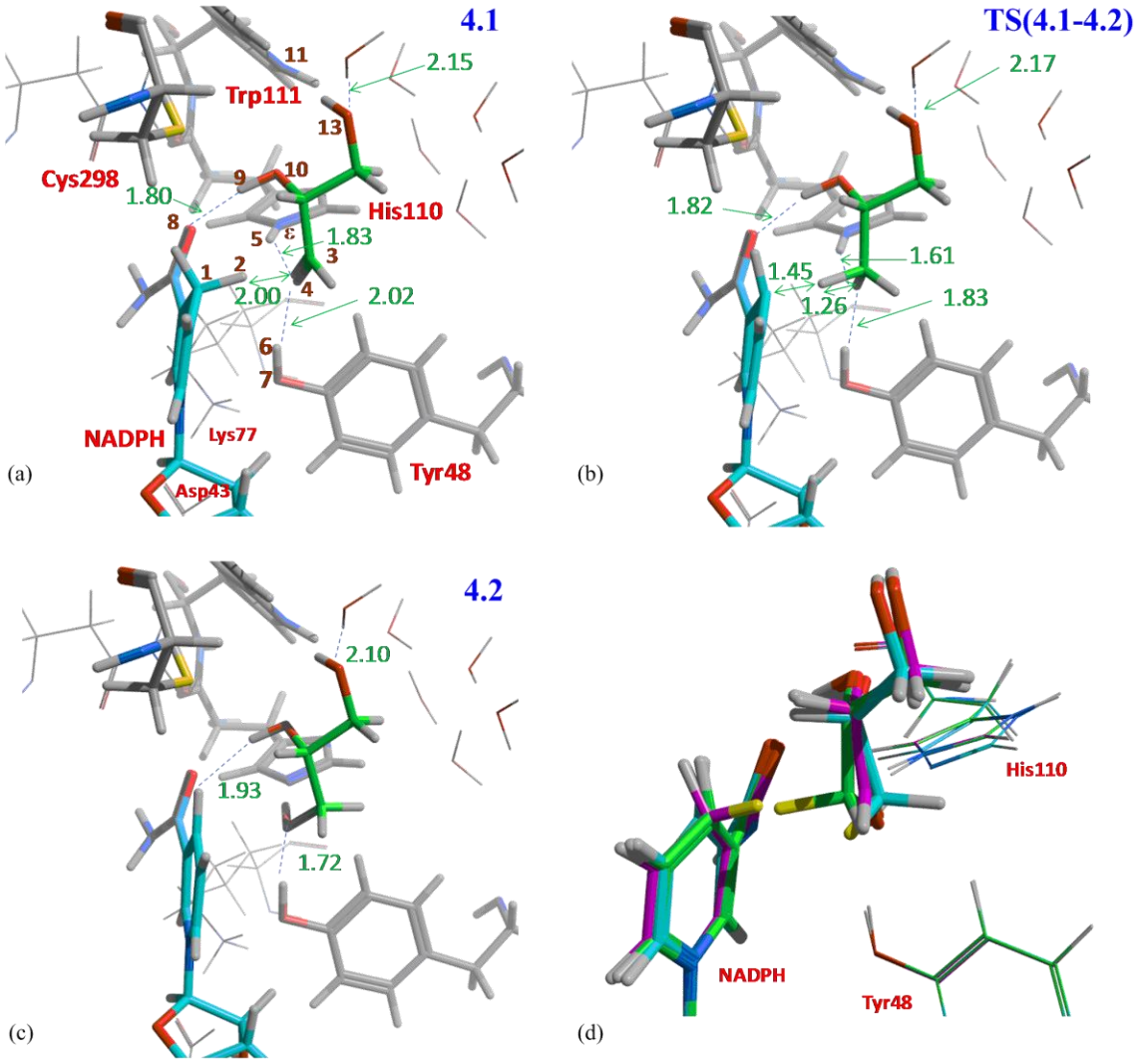


Figure 4.4. (a-c) Reaction intermediates of the GLD reduction by AR with HIP110 as studied by the QM/MM model (a) Enzyme-substrate complex (b) Transition state (c) Enzyme-product complex (Distances shown in green, atoms numbers in brown and residues names in red). (d) Superposition of the three intermediates of the reaction (**4.1** in pink, **TS(4.1-4.2)** in green, **4.2** in cyan and hydride in yellow).

Significant interactions help to maintain atoms in the ES complex, **4.1**, in a suitable position for reactions. Namely, H-bonds between both His110 and Tyr48 hydroxyl group

and the carbonyl group of GLD (H5(His110)···O4(GLD) = 1.83 Å and H6(Tyr48)···O4(GLD) = 2.02 Å), the H-bond between the NADPH amide group and the 2-hydroxy of GLD (O8(NADPH)···H9(GLD) = 1.80 Å), and finally the H-bond between the 3-hydroxy of GLD with a water molecule (H(H_2O)···O13(GLD) = 2.15 Å). In the transition state (characterized by an imaginary frequency of -667 cm^{-1}) the NADPH hydride is approximately halfway between C1 and C3, the C1···H2 and C3···H2 distances being 1.45 and 1.26 Å, respectively (**Figure 4.4b**). In addition, the comparison of the enzyme-substrate complex (**Figure 4.4.a**) and TS (**Figure 4.4b**) geometries shows the beginning of transition from a planar sp^2 to a tetrahedral sp^3 for the GLD carboxyl. In the same way, the donation of the hydride by NADPH makes the nicotinamide ring become more planar. The TS structure **TS(4.1-4.2)** does not clearly show whether His110 or Tyr48 is the proton donor: both H5 from His110 and H6 from Tyr48 are now closer to O4 (GLD) (H5(His110)···O4(GLD) distance is 1.61 Å and H6(Tyr48)···O4(GLD) distance is 1.83 Å) and these may contribute to the stabilization of the TS. However, from the product complex (**Figure 4.4c**) it is clear that the proton donor is His110, thus the role of Tyr48 is stabilizing the incipient negative charge on the aldehyde group of GLD. From these results it can be concluded that the mechanism for the protonated histidine system is concerted and asynchronous, where the approach of the NADPH hydride to the carbonyl carbon of GLD triggers the proton transfer. In the final product complex (EP), the hydride is definitively bonded to the D-glycerol carbon C3, and the proton H4 from His110 has been completely transferred to D-glycerol oxygen O4. A strong hydrogen bond involving H6 of Tyr48 persists, the O4···H6 distance being 1.72 Å.

We examined the possibility of an alternative mechanism where the proton transfer (from H5(His110) to O4(GLD)) and the attack of the hydride on the C3 group occur in two separate steps. In spite of an extensive search it was not possible to locate any intermediate corresponding to the alkoxide, and thus the possibility of a two-step non-concerted mechanism was discounted. Finally, the mechanism with a proton transfer from Tyr48 was also intensively investigated, but no transition state that could lead to a proton transfer from Tyr48 could be identified.

It is also informative to superpose the three stationary points along the potential energy surface (**Figure 4.4d**). The aldehyde hydrogen of the reactant maintains its location in the TS, but is replaced by the NADPH hydride hydrogen at the product stage. This suggests that the active site is set up to stabilize a hydrogen at this point and is ideally arranged for this transformation.

4.3.2 Mechanism with HIE110

For HIE, the mechanism is constituted of two steps with an activation energy of 16.0 kcal/mol (**Scheme 4.3b** and **Figure 4.3**). A schematic representation of the starting enzyme-substrate complex **4.3**, the two transition states (**TS(4.3-4.4)** and **TS(4.4-4.5)**), the intermediate **4.4** and the final enzyme-product complex **4.5**, including only the closest atoms around substrate, is given in **Figure 4.5a-e**.

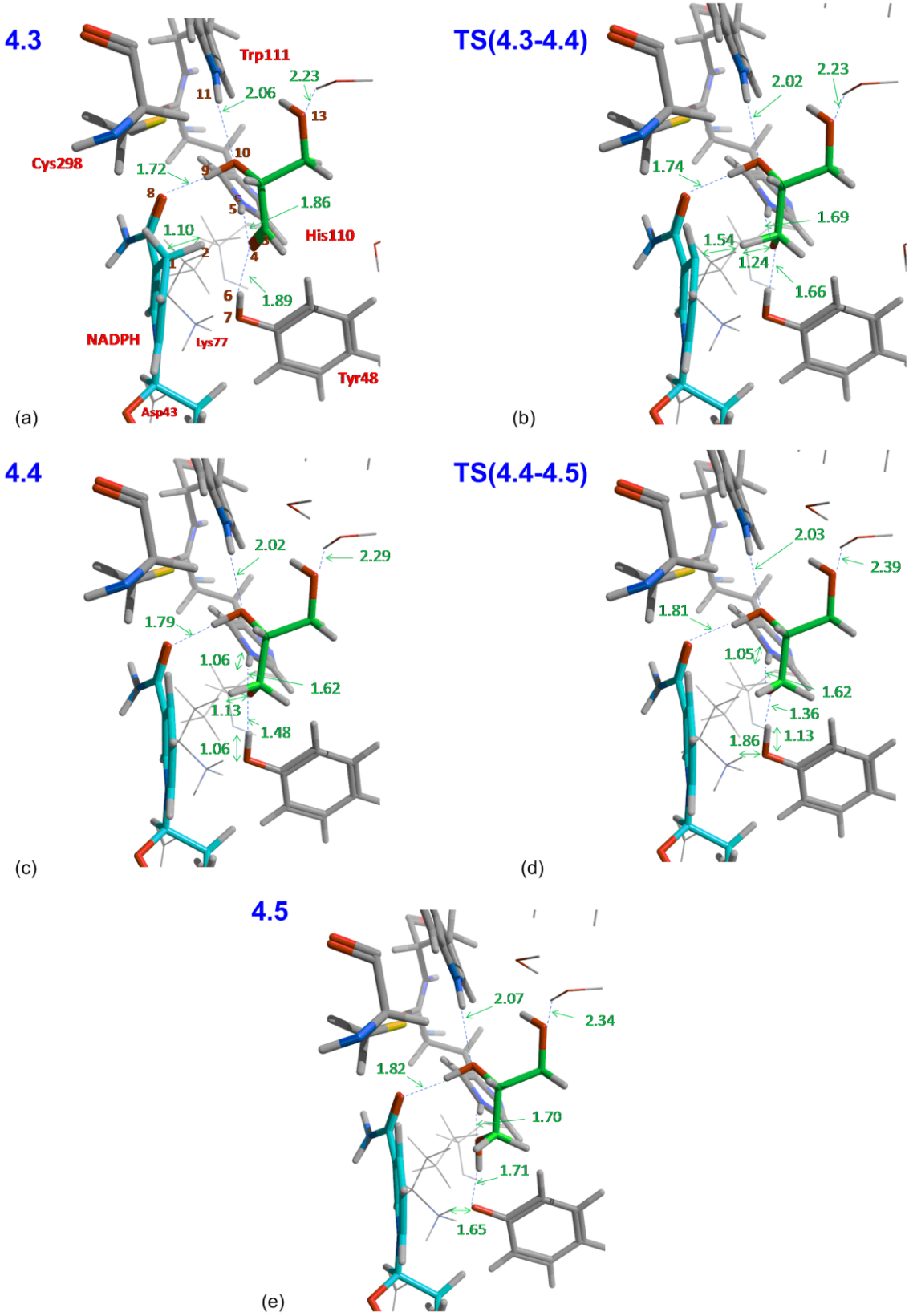

4.3
Trp111
2.23
11
2.06
13
Cys298
10
1.72
9
8
5
1.86
1.10
His110
2
1
4
6
1.89
7
NADPH
Lys77
Tyr48
Asp43
(a)
TS(4.3-4.4)
2.02
2.23
1.74
1.69
1.54
1.24
1.66
(b)
4.4
2.02
2.29
1.79
1.06
1.62
1.13
1.48
1.06
(c)
TS(4.4-4.5)
2.03
2.39
1.81
1.05
1.62
1.36
1.86
1.13
(d)
4.5
2.07
2.34
1.82
1.70
1.71
1.65
(e)

Figure 4.5. Reaction intermediates of the GLD reduction by AR with HIE110 as studied by the QM/MM model. (a) Enzyme-substrate complex (b) Transition state 1 (c) Enzyme-intermediate complex (d) Transition state 2 (e) Enzyme-product complex (Distances shown in green, atoms numbers in brown and residues names in red).

The same interactions that help to maintain atoms in the ES complex **4.1** can be found in **4.3**. These are hydrogen bond interactions between both His110 and Tyr48 residues and the carbonyl oxygen of GLD (H5(His110)···O4(GLD) = 1.86 Å and H6(Tyr48)···O4(GLD) = 1.89 Å) and the interaction between the amide group of NADPH and the middle hydroxyl group of GLD (O8(NADPH)···H9 = 1.72 Å) (**Figure 4.5a**). A water molecule also stabilizes the 3-hydroxy of GLD (O13(GLD)···H2O = 2.23 Å). Compared to **4.1**, **4.3** is further stabilized by a supplementary interaction with Trp111 (H11(Trp111)···O10(GLD) = 2.06 Å) that is not always present in the HIP simulation.

In the transition state **TS(4.3-4.4)**, (characterized by an imaginary frequency of -562 cm^{-1}) the NADPH hydride transfer from C1 to C3 is nearly completed, the C1···H2 and C3···H2 distances being 1.54 Å and 1.24 Å, respectively (**Figure 4.5b**). Hydrogens from Tyr48 and His110 are both almost at the same distance to the carboxyl oxygen of GLD O4 and closer compared to **4.3** – the H5(His110)···O4(GLD) distance was 1.86 Å in **4.3**, but is 1.69 Å in **TS(4.3-4.4)** and H6(Tyr48)···O4(GLD) distance was 1.89 Å in **4.3** and is 1.66 Å in **TS(4.3-4.4)**. The H11(Trp111) to O10(GLD) distance is nearly unchanged from the **4.3** (2.06 Å) to **TS(4.3-4.4)** (2.02 Å). The interaction between the amide group of NADPH and the 2-hydroxy of GLD (O8(NADPH)···H9(GLD) = 1.74 Å) is relatively unchanged at **TS(4.3-4.4)** compared to **4.3**, suggesting that the function of this interaction

is to maintain the position of the substrate through a consistently strong stabilizing interaction.

In the intermediate **4.4**, the NADPH hydride is now completely transferred from C1 to C3 as C3···H2 is 1.13 Å, the distance of a C-H bond. The difference between His110 and Tyr48 is clear in the intermediate structure as Tyr48 H6 is closer to the GLD O4 (1.48 Å) than His110 H5 is to O4 (1.62 Å). The transition from **4.4** to **TS(4.4-4.5)** is almost barrier less with a difference of 0.4 kcal/mol.

In the second transition state **TS(4.4-4.5)**, (characterized by an imaginary frequency of -305 cm^{-1}) the hydrogen bonding from Tyr48 and His110 to the GLD carbonyl oxygen O4 is further differentiated. This proton is partially transferred from Tyr48 (H6(Tyr48)···O4(GLD) =1.36 Å) compared to His110 in which the hydrogen bonding remains consistent relative to **4.4** (1.62 Å).

In the final EP complex **4.5**, the proton H6 from Tyr48 is bonded to O4 from GLD. At this stage a strong interaction is formed between the formed Tyr48 phenolate and Lys77, going from 1.86 Å in **TS(4.4-4.5)** to 1.65 Å in **4.5**.

4.3.3 Mechanism with HID110

During the HID110 1 ns MD simulation a displacement of NADPH occurred (**Figure 4.6a**), giving an unproductive complex (with the NADPH hydride too far from the carbonyl carbon of the ligand) and perhaps suggesting that this electronic state is quite

unreactive. To address this issue, the energy profile for HID was determined with a new 1 ns MD, but with restraints on the co-factor position.

For HID the eventually identified mechanism is a concerted one, using Tyr48 as proton donor with an activation energy of 26 kcal/mol (**Scheme 4.3c** and **Figure 4.3**). A schematic representation of the starting enzyme-substrate complex **4.6**, the transition state **TS(4.6-4.7)** and the final enzyme–product complex **4.7**, including only the closest atoms around substrate, is given in **Figure 4.6b-d**.

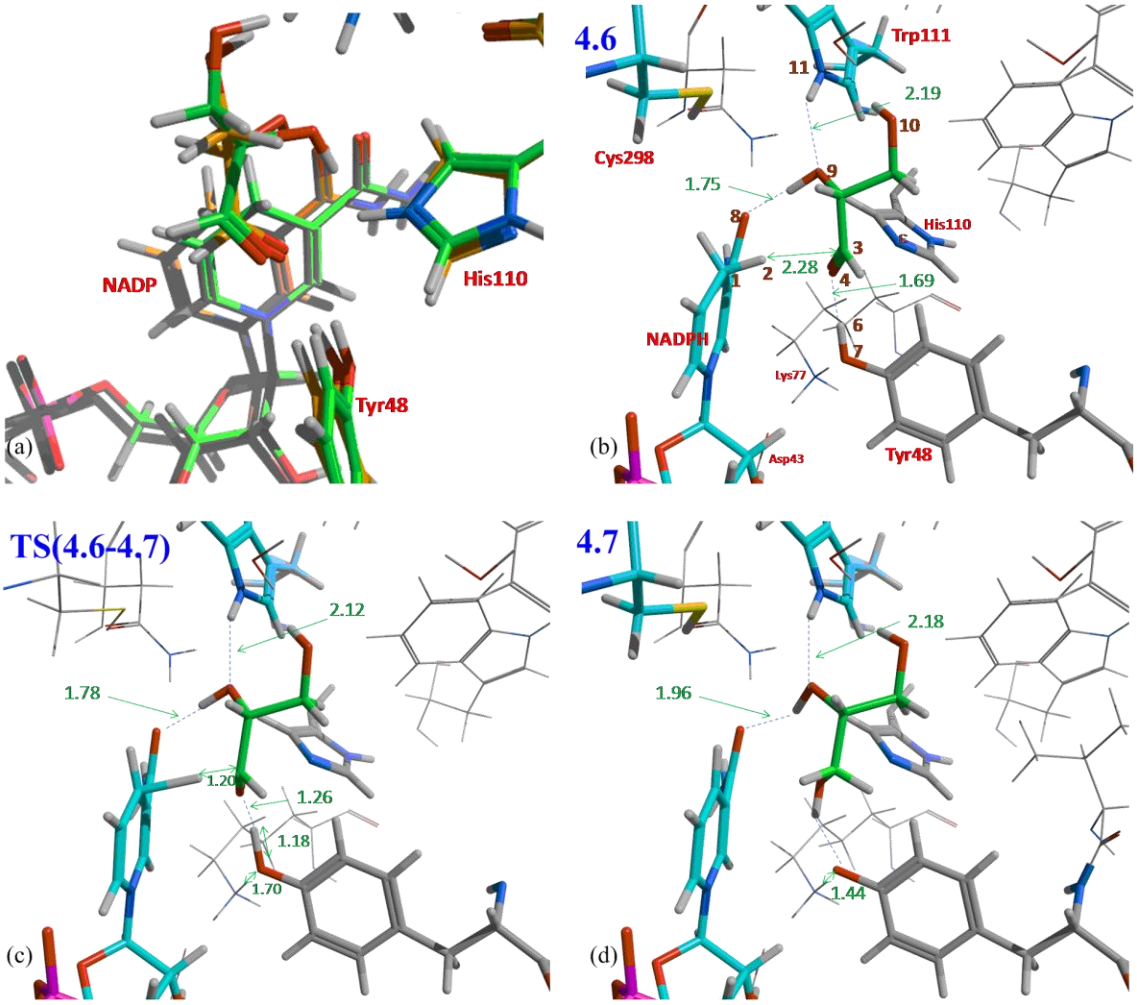


Figure 4.6. (a) Comparison of NADPH position between HID (green) and HIE (orange) after a 1ns MD simulation. (b-d) Reaction intermediates of the D-glyceraldehyde reduction by AR with HID110 as studied by the QM/MM model (b) Enzyme-substrate complex (c) Transition state (d) Enzyme-product complex (Distances shown in green, atoms numbers in brown and residues names in red).

In the case of HID there are fewer interactions that help to maintain atoms in the ES complex **4.6**, in a suitable position for reactions. As Nε of His110 is deprotonated, no stabilization is possible. In contrast to the other simulations, no water was observed interacting with the 3-hydroxy of GLD. As a consequence the interaction between the Tyr48 hydroxyl group and O4 of GLD carbonyl group is strong (H6(Tyr48)···O4(GLD) = 1.69 Å) whereas in **4.1** and **4.3** the hydrogen bond was longer, 2.02 Å and 1.89 Å, respectively (*cf.* **Figure 4.4a**, **Figure 4.5a** and **Figure 4.6b**). The NADPH hydride is almost at the same distance from the GLD carboxyl carbon (H2(NADPH)···O4(GLD) = 2.28 Å) compared to **4.3** (2.25 Å) but slightly further compared to **4.1** (2.00 Å). The hydrogen bond between the GLD 2-hydroxy and Trp111 is fairly consistent between both **4.6** (2.19 Å) and **4.3** (2.20 Å).

In the transition state **TS(4.6-4.7)** (characterized by an imaginary frequency of -875 cm^{-1}), the NADPH hydride is moving from C1(NADPH) to C3(GLD). The hydride is very close to completely transferred as the distance to C3 is only 1.20 Å. The proton from Tyr48 is approximately halfway between O7(Tyr48) and O4(GLD) (H6···O7(Tyr48) distance is 1.18 Å and H6(Tyr48)···O4(GLD) distance is 1.26 Å.) Thus, the mechanism

is concerted, indeed almost simultaneous, between the hydride transfer and the proton transfer.

In the final EP complex **4.7**, the hydride is definitively bonded to D-glycerol carbon C3, and the proton H6 from Tyr48 has been completely transferred to GLD O4. At this stage a strong H-bonding interaction is formed between the phenolate of Tyr48 and Lys77 (1.44 Å).

4.4 Comparison between the three reaction models

The calculated activation barriers for HIP and HIE are of 8.1 kcal/mol and 16.0 kcal/mol, respectively. Thus from an energetic point of view both mechanisms are different. This is all the more true when we average the results obtained from QM/MM studies on other frames (one additional for HIP and two for HIE) that we also studied (Appendix Figures A.1.2, A.1.3 and A.1.4). These gave an average of 6.5 +/-2.2 kcal/mol for HIP and 16.7 +/-1.0 kcal/mol for HIE. Also, it should be pointed out that we are comparing ΔE with ΔG, nevertheless in these types of reactions the contribution from thermal and entropic effects is expected to be small.[72] It could be concluded from these results that the mechanism with the lower activation energy is the more probable one. Nevertheless, the experimental activation free energy calculated from kinetics constants is 14.8 kcal/mol.[298] Thus, although the activation energy with the HIP model is lower, the results from the HIE model are closer to the experimental value. Given the significant difference (~ 10 kcal/mol) between the calculated activation energies that arises from considering the protonation state of the histidine – the clear agreement of the HIE model with the experimental data indicates that the experimental system involves an unprotonated histidine in the binding site with Tyr48 acting as the proton donor. These conclusions based on the difference in the calculated activation barriers are agreement with the calculation of the pK_a of the residues, which has been done by Várnai and Warshel,[296] that yielded an estimated pK_a of 8.5 for Tyr48 and a remarkably low value of 0.9 for His110. The study of the HID model gave a much higher energy barrier of 26.5 kcal/mol. The

difference in the results of HIE and HID models demonstrate that the presence of a proton on Nε of His110 is required for the correct positioning of GLD.

4.5 Effect of basis set size and QM region size

The goal of this work was to obtain an updated QM/MM model of the reduction of GLD by AR to determine both the mechanism of reaction and the effect that a different QM/MM methodology can have on the outcome of results. To reach that goal we have used a more accurate QM treatment and a larger QM zone. Thus, in the following, the current results are compared with those from previous studies.

Our HIP model can be compared to the results obtained by Lee and co-workers[295] and our HIE model to the results obtained by Várnai and co-workers.[296] To help the comparison, energies from both studies and this work have been summarized in **Table 4.2**. The structural characteristics were very similar to previous studies and are thus described in the appendix Table A1.1.

Table 4.2. Comparison of previous QM/MM results from Lee and co-workers[295] model (Lee) and Várnai and co-workers[296] model (Var.) to HIP and HIE models.[a,b,c]

Study	**Level of theory**	**TS1**	**I**	**TS2**	**P**
		proton donor His110			
		TS(4.1-4.2)			**4.2**
This work (HIP model)	B3LYP/6-31G*	8.1	/	/	-17.7
		Lee-TS1			Lee-P
Lee *et al.*[295]	HF/4-31G	21.2	/	/	-12.4
		proton donor Tyr48			
		TS(4.3-4.4)	**4.4**	**TS(4.4-4.5)**	**4.5**
This work (HIE model)	B3LYP/6-31G*	16.0	12.4	12.8	3.9
		Var.-TS1	Var.-I	Var.-TS2	Var.-P
Várnai *et al.*[296]	AM1	41.2	33.6	34.7	10.3

[a] TS: transition state, I: intermediate and P: product

[b] Electronic energies (ΔE) in kcal/mol are given relative to the reactant state for each system studied

[c] The QM/MM partitioning employed in Lee *et al.* study and in this work are described in Figure 4.2a and 4.2b, respectively. In Várnai *et al.* study the QM region is larger than in the Lee *et al*. study but smaller than in this work (the same residues are included in the QM region but the link atom positioning is different).

From an energetic point of view the results differ significantly. For the HIE model there is notable difference in the activation energy between the work of Várnai and co-

workers[296] [41.2 kcal/mol] and this work [16.0 kcal/mol], a difference of 25.2 kcal/mol. From the experimental activation free energy calculated from kinetics constants of 14.8 kcal/mol we know that our model is in better agreement.[298]

For the HIP model, the individual influence of the QM treatment and the QM size is summarized in **Table 4.3**.

Table 4.3. Analysis of the effect of the QM treatment and QM size on the activation energy by comparing results from this work (HIP model) and previous work from Lee and co-workers[295].[a]

Study	**Lee *et al.*[295]**	**This work (HIP model)**	
QM treatment	HF/4-31G	B3LYP/6-31G*	B3LYP/6-31G*
QM/MM partitioning[b]	1	1	2
E_a	21.2[c]	7.5	8.1
ΔE	-12.4[c]	-6.0	-17.7

[a] Electronic activation energies (E_a) and reaction energies (ΔE) in kcal/mol are given relative to the reactant state for each system studied.

[b] as defined in Figure 4.2

[c] reference 295

In the Lee and co-workers study the relative energy to the reactant of the TS for the His110 proton donor model, obtained using 4-31G, was 21.2 kcal/mol.[295] From our results we can see that the use of the more accurate B3LYP/6-31G* method has significantly changed the calculated relative energies as we obtained an activation energy of 7.5 kcal/mol, 13.7

kcal/mol smaller than the HF/4-31G method. The combination of B3LYP/6-31G* and a larger QM region did not significantly alter the activation energy (7.5 kcal/mol for the smaller region and 8.1 kcal/mol for the larger). However, the effect on the relative energy of the product to the reactant (-6.0 kcal/mol for the smaller region and -17.7 for the larger) was more substantial. Overall, the results show a meaningful gain in accuracy for the comparison of the two potential reaction mechanisms is due mostly to the developments in accuracy and efficiency of QM methods.

4.6 Conclusion

Since 1992 and the first suggestion of His110 and Tyr48 as potential proton donors, there has been a long history of debate on the catalytic mechanism of AR.[291, 308] Nevertheless the common opinion seemed to favor the Tyr48 proton donor mainly because of crystallographic and mutagenesis data.[294, 313-316]

Nevertheless two previous QM/MM methodologies (using HF 4-31G/CHARMM22[245] and AM1[312]/CHARMM22[245]) have failed to validate the Tyr48 hypothesis. Furthermore they have also given different results between them for the proposed mechanism; one predicted a concerted mechanism while the other predicted a step-wise mechanism. By using a different force field and QM method (OPLS2005[243] and B3[277]LYP[234, 278]/6-31G*[279]) and a bigger QM region, the mechanism was re-evaluated. A different mechanism is suggested depending on the protonation state of His110. With HIP as protonation state for His110, the results show an average activation energy of 6.5 ± 2.2 kcal/mol and evidence for a highly asynchronous concerted mechanism with His110 as proton donor. With HIE, the mechanism is different, as results show an average activation energy of 16.7 ± 1.0 kcal/mol and evidence for a step-wise mechanism using Tyr48 as proton donor. Preliminary MD simulation on HID indicates that this protonation state is unreactive and shows the importance of a proton on Nε of His110 for the reaction to occur as this residue is implicated in the positioning of the substrate prior to the reaction. Our results demonstrate that the HIP and HIE model mechanisms are significantly different in energy and that only the HIE model is in good agreement with experimental data – confirming that Tyr48 the most probable proton donor. Finally the effect of using modern

DFT methods for the QM/MM calculation was evaluated by comparing our results to previous studies. We found that the changes in energetics can be substantially affected by the choice of methods and, importantly, the size of the QM site (particularly for the relative energy of the reactants and products).

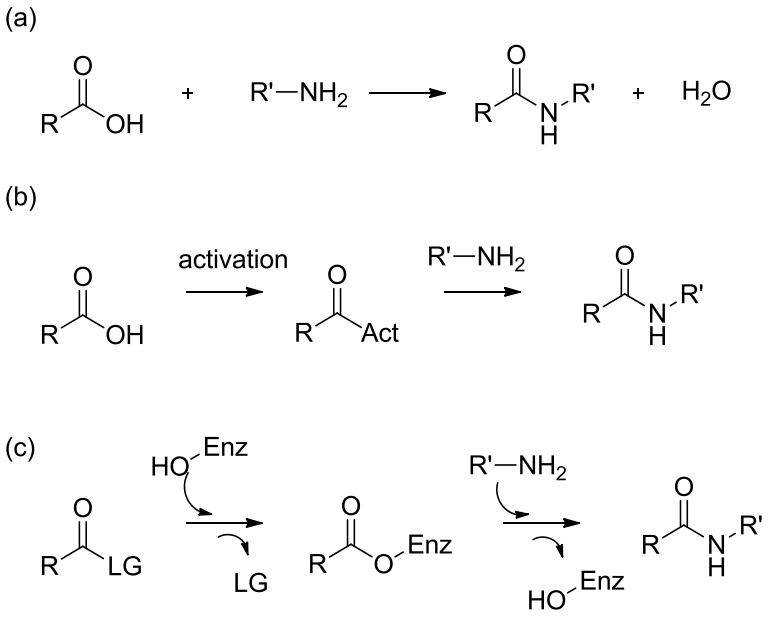
The work described in this chapter has addressed our first aim – establish a validated internal QM/MM methodology. This validated methodology has been used as starting point for the elaboration of a QM/MM-based hotspot identification protocol which is described in the next chapter.

5. Computational investigation of a carboxylesterase catalyzed amide bond formation reaction

5.1 Introduction

5.1.1 Amide bond formation

"Amide bond formation avoiding poor atom economy regents" was voted the highest priority area of research in 2006 by the American Chemical Society's (ACS) Green Chemistry Institute (GCI) Pharmaceutical Roundtables (PR).[5] While a lot of research has been done since then to overcome this problem it remains a challenge today.[317] Amide bond formation is one of the most used reactions for the pursuit and preparation of drug candidate in the pharmaceutical industry.[318-319] These reactions are usually synthesized by the condensation of a carboxylic acid with an amine; ideally this would happen through direct condensation giving water as the only by-product (**Scheme 5.1a**).[320] In reality this is mostly done in several steps as the carboxylic acid needs to be activated prior to the reaction (**Scheme 5.1b**).[321] This activation usually requires a stoichiometric amount of coupling reagent generating a lot of waste. Thus, there is a necessity to develop alternative and more environmentally friendly processes. One alternative is the use of biocatalysts, which eliminates the need for a coupling reagent and gives good atom economy (**Scheme 5.1c**).



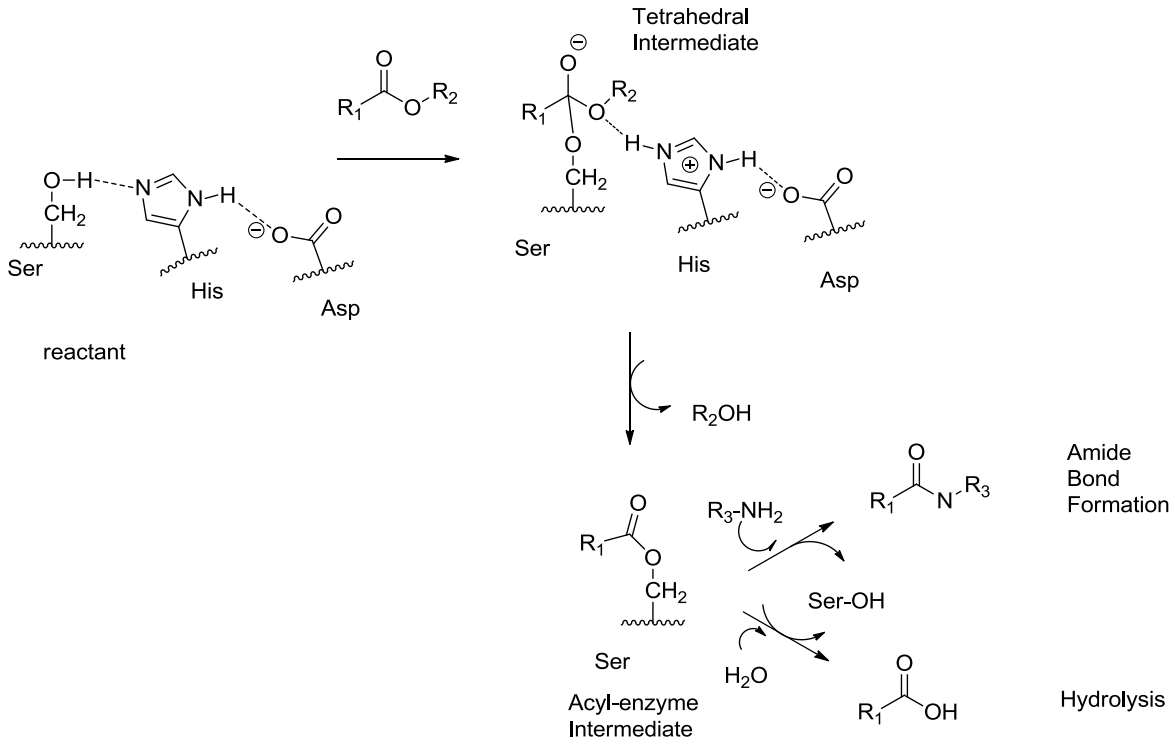
Scheme 5.1. Amide bond formation (a) Direct (b) Via an activation process (c) Enzymatic (Protease/Lipase/Esterase). LG = leaving group; Enz = enzyme.

5.1.2 Esterases

Carboxylester hydrolases (EC 3.1.1), commonly named esterases, consist of a large spectrum of enzymes defined by their ability to catalyze the hydrolysis of carboxylic ester bonds and are widely distributed among animals, plants, and microorganisms.[322] Historically, members of carboxylester hydrolases group have been classified into two major subclasses on the basis of their known substrate specificity: 'true' esterases (carboxylesterases; EC 3.1.1.1) and lipases (triacylglycerol hydrolases; EC 3.1.1.3). However, lipases and esterases consensus motifs described by the ProSite[323] database are very close and are thus usually very difficult to differentiate from the structure.[322] Therefore, a classification was made depending on the substrate preference; indeed

esterases usually hydrolyse carboxyl esters of short-chain acylglycerol (≤10 carbon atoms), while lipases prefer carboxyl esters of long-chain acylglycerol (≥10 carbon atoms).[324]

Among enzymes, carboxylester hydrolases are particularly interesting biocatalysts for industrial application as several of them have proven to be stable under a variety of conditions, while working without organic cofactors.[325-326] In nature, esterases' primary function is to cleave carboxyesters (RCOOR′) into the corresponding carboxylic acid (RCOOH) and alcohol (R′OH) via a proton transfer hydrolysis mechanism using a catalytic serine present within a Ser-His-Asp catalytic triad and through an acyl-enzyme intermediate (**Scheme 5.2**). Amide bond formation can also happen from the results of the interception of the formed acyl-enzyme intermediate by N-nucleophiles such as ammonia or amines.[325, 327] To favor this amidation reaction particular reaction conditions are needed, such as water removal and use of organic solvents, which pushes the reaction equilibrium to the side of the product and avoids the reverse hydrolysis reaction.[327]

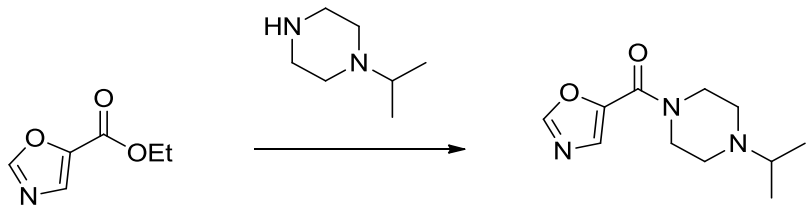


Scheme 5.2. Mechanism for an esterase.[325]

5.1.3 Target reaction

Within GlaxoSmithKline an enzymatic option in the synthetic route of a recently disclosed Phosphoinositide 3-Kinase δ inhibitor for the treatment of respiratory disease has been investigated (**Scheme 5.3**).[328-329] From the patent it can be seen that this reaction represents the first two steps of the synthesis. From the initial screen an esterase was retained for further evolution of the process. This was a thermophilic carboxylesterase (E.C 3.1.1.1) from *Alicyclobacillus acidocaldarius* (K. Brown, GSK, personal communication). There are four X-ray crystallographic structures available in the Protein Data Bank (PDB) for this enzyme, one wild type (PDB ID: 1EVQ) and three mutant

structures (PDB ID: 2HM7, 1QZ3, 1U4N). The enzyme has a classic hydrolase family fold with 8 highly twisted central β-strands surrounded by nine α-helices (**Figure 5.1a**). The active site is composed of the classic catalytic triad with residues Ser155, His282 and Asp252 and an oxyanion hole formed by the backbone of residues Gly83, Gly84 and Ala156 (**Figure 5.1b**).



Scheme 5.3. Target reaction for enzymatic transformation.

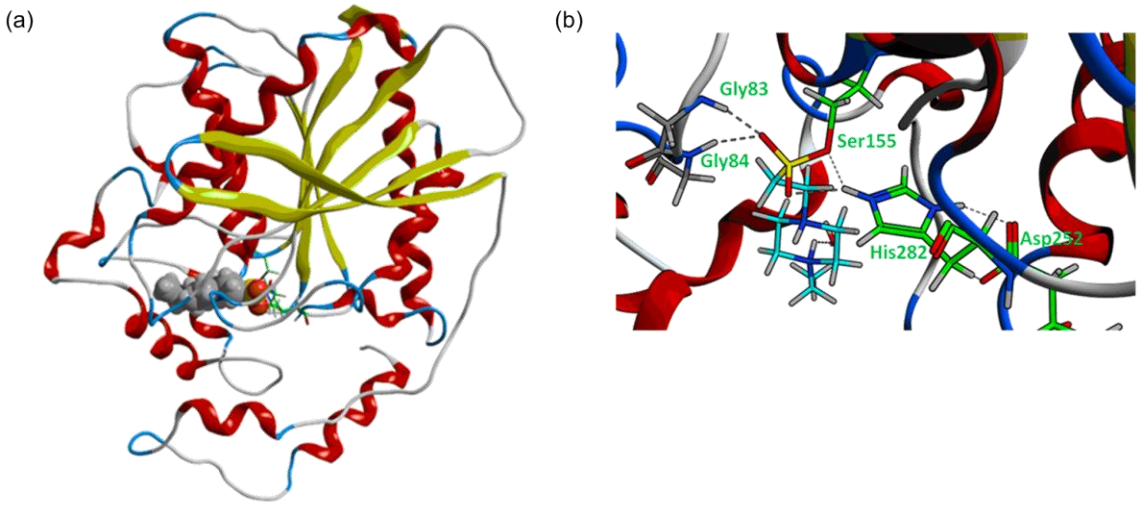


Figure 5.1. (a) Crystal structure of carboxylesterase from *Alicyclobacillus acidocaldarius* (PDB ID: 1EVQ) (b) Active site view with Hepes covalently bounded.

In the present work, we have first examined the catalytic mechanism of this enzymatic amide bond formation using the QM/MM methodology established in **Chapter 4**. This information has then been used as a starting point for the charge modification procedure 'ElectroScan', as introduced in Section 3.3.7, with the objective of identifying hotspot residues. Possible limitations of the 'ElectroScan' are listed and possible ways to overcome them are discussed through the use of 'PosiScan', another procedure introduce in Section 3.3.7.

5.2 Computational methods

5.2.1 Model systems

Theoretical studies were performed starting with an X-ray crystal structure (PDB ID : 1EVQ) of the carboxylesterase enzyme that has a resolution of 2.6 Å.[330] The structure comprises a Hepes molecule ((4-hydroxyethyl)-1-piperazine ethane sulphonic acid) covalently bound to Ser155, so this was transformed into a reactant configuration by deleting this Hepes molecule to obtain a free Ser155. The first ligand of the reaction, ethyl oxazole-5-carboxylate (EOC), was then build in place of the Hepes molecule. The carbonyl carbon of EOC was positioned in a configuration that would enable a nucleophilic attack of Ser155. In order to get an adequate pose, a conformational search with constraints on the distances between the carbonyl carbon of EOC and the oxygen of Ser155 was run. The parameters for EOC were taken from the OPLS3[331] force field. The system was prepared and solved as described in Section 3.3.4. Finally, to neutralize the system, 15 sodium atoms were added to the system in random positions.

The model system was gradually relaxed using a standard protocol implemented in Desmond.[270-271] A molecular dynamics (MD) simulation was performed for 1.2 ns using Desmond[270-271] in order to relax further the system. Details of the MD simulations are described in Section 3.3.4 of the method chapter. The quality of the simulation was checked by plotting the different RMSD which results can be found in **Appendix Figure A.2.1**.

One snapshot was extracted from the MD output and prepared for QM/MM calculations. The selected snapshot was chosen on the basis of geometric parameters, more specifically the shorter distance between the carbonyl carbon of EOC and the oxygen of Ser155 to enable a nucleophilic attack. While the more rigorous approach is to calculate the reaction energetics from multiple snapshots, in the case of screening for residues to mutate a single snapshot is sufficient to determine the relative effect of the each residue in the shortest possible time.[332] The selected snapshot was MM minimized by using the truncated Newton method[265] implemented in Impact[306]. This post-equilibration, minimized structure, which represents the enzyme-substrate (ES) complex, was used as the starting structure for the QM/MM calculations.

5.2.2 QM/MM Methodology

The QM region for the study includes all hypothetical reacting species (Ligand, Ser115 and His282) and stabilizing residues (Gly83, Gly84, Ala156 and Asp252). This QM region is represented in **Figure 5.2**, it contains 66 atoms in the acylation step and 82 in the amidation step due to the difference of ligand present in the active site.

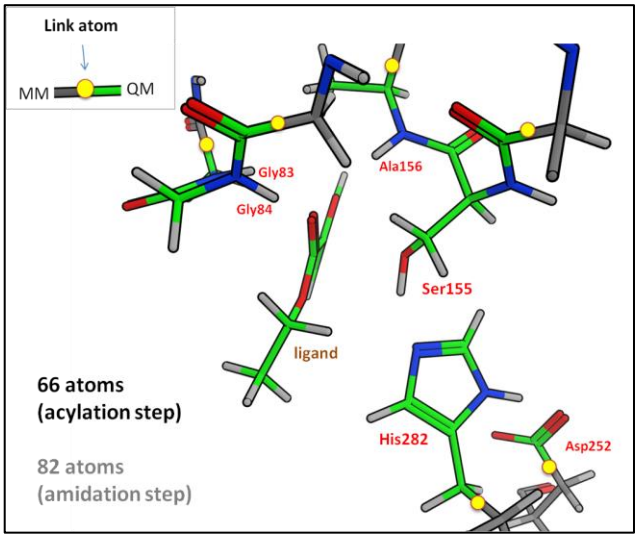


Figure 5.2. QM/MM partitioning used for the calculations (acylation step).

All atoms beyond 15 Å from the reactant were kept constrained during the QM/MM simulations in order to speed up the calculations. The equilibrated ES complex was optimized with QM/MM calculations using B3[277]LYP[234, 278]/6-31G*[279] level of theory for the QM region and the OPLS2005[243] force field for the classical region (MM). The potential energy surface (PES) for the reaction was explored starting from this optimized structure of the reactant as described in the method chapter Section 3.3.6.

To evaluate the electrostatic impact of each residue of the protein on the reaction barrier the 'ElectroScan' and 'PosiScan' procedures, as described in Section 3.3.7, were applied.

5.3 Proposed reaction mechanism for the carboxylesterase amide bond formation

In this work, the catalytic mechanism of an amide bond formation catalyzed by a carboxylesterase is unravelled. The reaction follows a two-step mechanism: first acylation and then amidation through a tetrahedral intermediate. In the following section these two steps will be described in two separate sections.

5.3.1 Formation of the Acyl–Enzyme Complex

The results for the formation of the acyl–enzyme complex requires a two-step mechanism: first the nucleophilic attack of Ser155 to the substrate and second the concomitant formation of the first product of the reaction, an alcohol molecule (**Figure 5.3a**). A schematic representation of the two transition states of the reaction (**TS(5.1-5.2)** and **TS(5.2-5.3)**) including only the closest atoms around the substrate, is given in **Figure 5.3b-c**. **Table 5.1** shows the relatives energies and key distances involved in the formation of the acyl–enzyme complex.

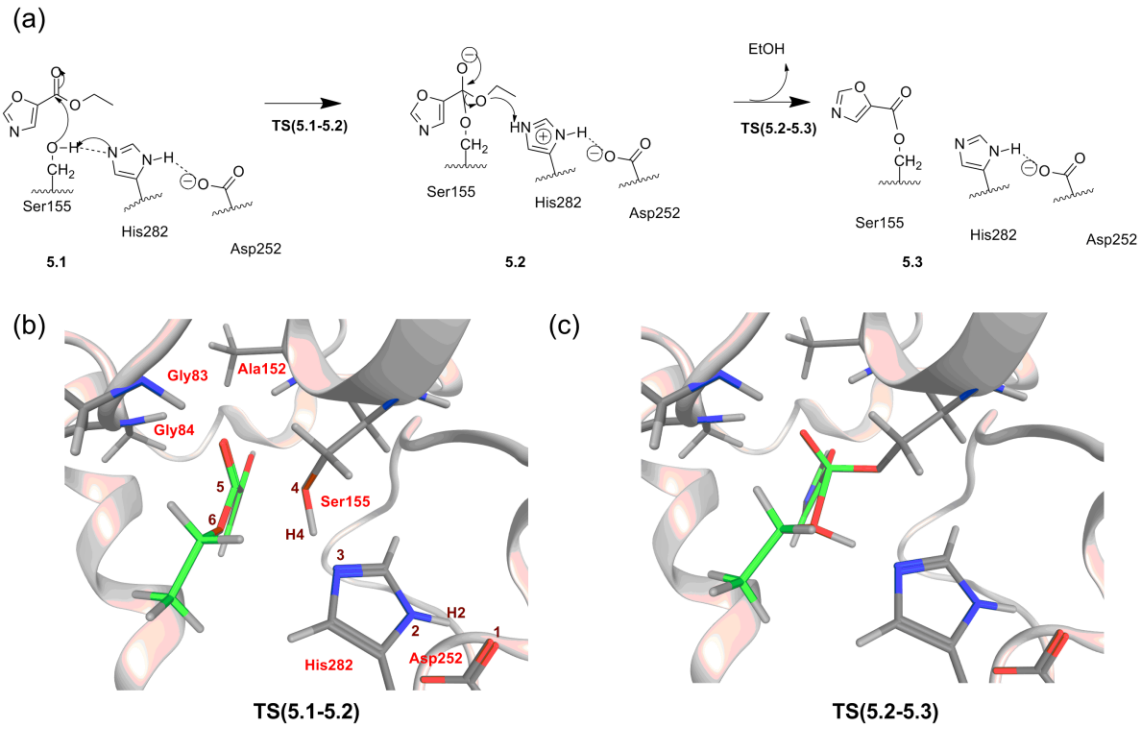


Figure 5.3. Schematic representation of the acylation pathway and the corresponding transition states structures.

Table 5.1. Relative energies and key distances during the formation of the acyl–enzyme complex.[a, b]

	ΔE (kcal/mol)	distances (Å)						
		O1···H2	**H2–N2**	**N3···H4**	**H4–O4**	**H4···O6**	**O4···C5**	**C5–O6**
5.1	0	1.64	1.05	1.62	1.01	2.66	2.46	1.33
TS(5.1-5.2)	5.2	1.57	1.06	1.21	1.25	2.51	1.93	1.36
5.2	-2.3	1.56	1.08	1.04	2.09	1.75	1.47	1.49
TS(5.2-5.3)	8.5	1.66	1.05	1.38	2.24	1.13	1.38	1.80
5.3	1.8	1.69	1.05	1.71	2.56	1.00	1.33	2.52

[a] Electronic energies (ΔE) in kcal/mol are given relative to the energy of **5.1**.

[b] Results obtained using B3LYP/6-31G*/OPLS2005 as level of theory

The first step of the acylation involves the nucleophilic attack of the oxygen of Ser155 on the first ligand of the reaction, ethyl oxazole-5-carboxylate (EOC). In the starting enzyme-substrate complex **5.1**, the carbonyl group of EOC establishes three hydrogen bond interactions: two of them with the backbone nitrogen atoms of Gly83 and Gly84 in the oxyanion hole region (1.83 and 1.98 Å respectively) and one with the backbone nitrogen atom of Ala156 (2.25 Å). Ser155 is very close to carbon C5 of the substrate (O4(Ser155)···C5(EOC) distance of 2.46 Å, atom numbering given in **Figure 5.3b**) and establishes a hydrogen bond with His282 (N3(His282)···H4(Ser155) distance of 1.62 Å). His282 also establishes another hydrogen bond with Asp252 (O1(Asp252)···H2(His282) distance of 1.64 Å). In the first transition state **TS(5.1-5.2)** (characterized by an imaginary frequency of -647 cm^{-1}) the proton that was bonded to Ser155 is now shared with His282

(H4(Ser155)···O4(Ser155) = 1.25 Å and N3(His282)···H4(Ser155) = 1.21 Å). All of this gives a stronger nucleophilic character to Ser155 that is closer to the C5 carbon of EOC (O4(Ser155)···C5(EOC) = 1.93 Å versus 2.46 Å in the reactants). In the first intermediate **5.2**, the proton that was previously bonded to Ser155 is now attached to His282, and Ser155 becomes covalently bound to the substrate at carbon C5 (O4(Ser155)–C5(EOC) bond length of 1.47 Å). This first intermediate of the reaction corresponds to a tetrahedral intermediate which is negatively charged but stabilized in the active site by the hydrogen bonds that are provided by the oxyanion pocket formed by residues Gly83, Gly84, and the Ala156 backbone. This first step requires an activation energy of 5.2 kcal/mol, and is mildly exothermic with a reaction energy of -2.3 kcal/mol (**Table 5.1**).

To obtain the acyl–enzyme complex a second step is required. In the second transition state **TS(5.2-5.3)** (characterized by an imaginary frequency of -398 cm^{-1}) the oxygen atom O6 is starting to dissociate from the tetrahedral intermediate (C5(EOC)···O6(EOC) distance of 1.80 versus 1.49 Å in the bound state) and the ε proton of His282 is being transferred to O6 (H4(His282)···O6(EOC) distance of 1.13 Å). In the product of this reaction **5.3** the acyl–enzyme is formed and an alcohol molecule is released. In the acyl-enzyme complex the distance between C5 and the oxygen atom of Ser155 is shorter (O4–C5 bond length of 1.33 Å) and the hydrogen bonds between the other oxygen atom of the substrate and the three backbone nitrogen atoms of the residues that form the oxyanion hole (Gly83, Gly84, and Ala156) remains present. This step requires an activation energy of 10.8 kcal/mol, and a change of the overall energy of reaction of 4.1 kcal/mol, relative the intermediate **5.2** (**Table 5.1**).

5.3.2 Amidation reaction

The investigation of the amidation reaction reveals that the amidation step requires a three-step mechanism: the nucleophilic attack of a second ligand amine nitrogen to the acyl-enzyme complex, the deprotonation of the nitrogen and finally the formation of an amide product and the regeneration of the active site (**Figure 5.4a**). A schematic representation of the three transition states of the reaction (**TS(5.4-5.5)**, **TS(5.5-5.6)** and **TS(5.6-5.7)**) including only the closest atoms around substrate, is given in **Figure 5.4b-d.** **Table 5.2** shows relatives energies and key distances during the amidation reaction.

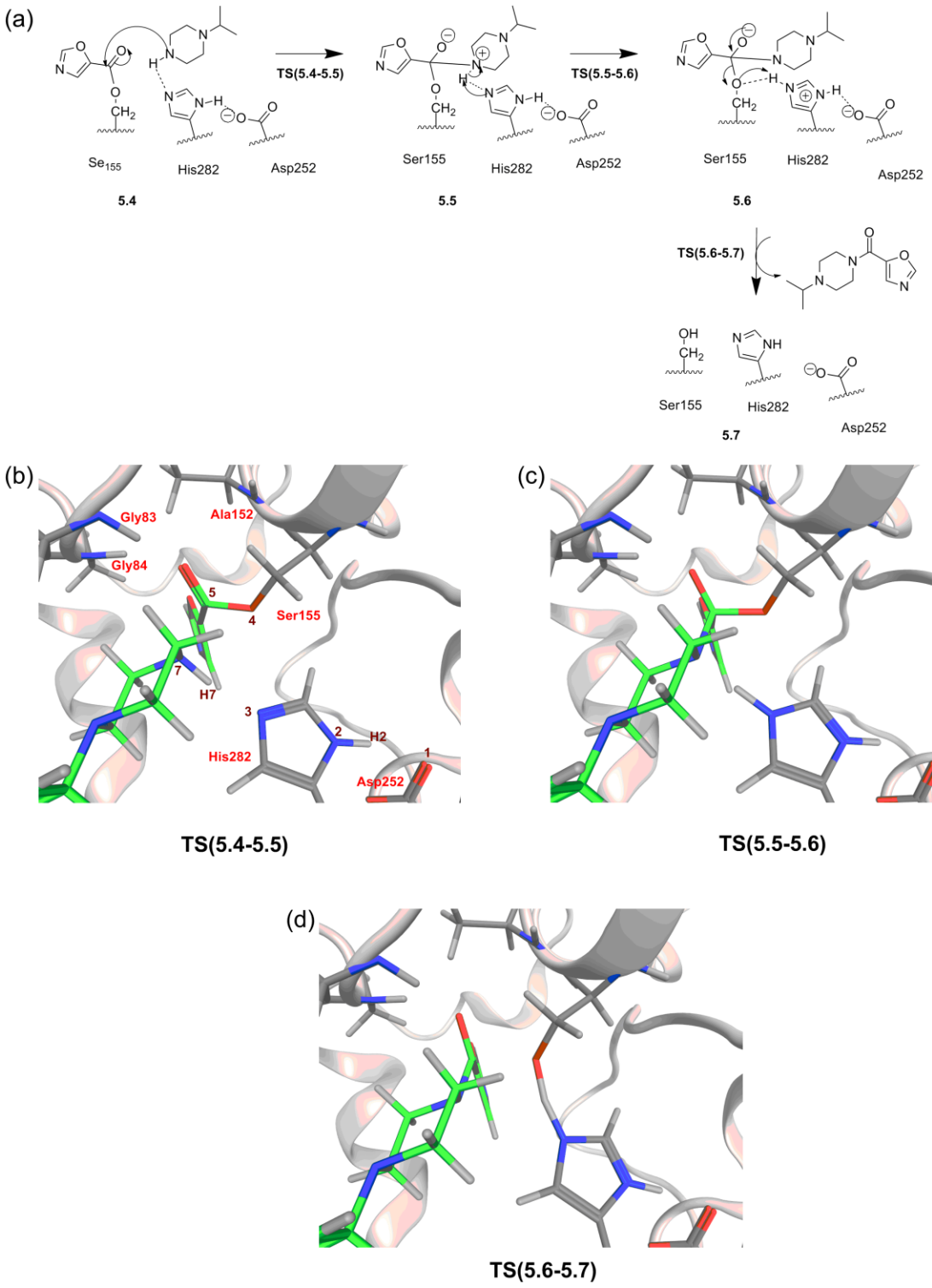


Figure 5.4. Schematic representation of the amidation pathway and the corresponding transition states structures.

Table 5.2. Relative energies and key distances during the amidation reaction.[a, b]

	ΔE (kcal/mol)	distances (Å)						
		O1···H2	**H2–N2**	**N3···H7**	**H7···O4**	**H7–N7**	**O4–C5**	**C5···N7**
5.4	0	1.92	1.04	2.15	3.06	1.03	1.34	3.61
TS(5.4-5.5)	17.3	1.86	1.04	1.74	2.51	1.04	1.39	2.72
5.5	17.1	1.87	1.04	1.70	2.40	1.06	1.43	1.76
TS(5.5-5.6)	23.7	1.85	1.05	1.17	2.38	1.45	1.50	1.59
5.6	23.5	1.80	1.05	1.12	2.36	1.53	1.51	1.56
TS(5.6-5.7)	28.3	1.66	1.05	1.22	1.23	2.42	1.84	1.46
5.7	12.7	1.78	1.04	1.69	1.00	3.07	2.75	1.38

[a] Electronic energies (ΔE) in kcal/mol are given relative to the energy of **5.4**.

[b] Results obtained using B3LYP/6-31G*/OPLS2005 as level of theory

Once the formation of the acyl–enzyme complex is completed and an amine molecule is available in the active site region, the amidation reaction can take place. In this reaction the alcohol group of the ester formed within the acyl–enzyme complex is exchanged to an amine group.

The first step of the amidation involves the attack of the nitrogen of the second ligand, 1-isopropylpiperazine (IPP), on the acyl-enzyme complex. In the optimized enzyme-substrate complex **5.4**, IPP is stabilized between the acyl complex and His282. One of the hydrogens of the amine group of IPP is pointing toward the Nε atom of His282 establishing a hydrogen bond (N3(IPP)···H7(IPP) distance of 2.15 Å). Such an interaction

allows the approach of the nitrogen atom near carbon C5 of the acyl–enzyme complex (C5(EOC)···N7(IPP) distance of 3.61 Å), promoting the amidation process. In the first transition state **TS(5.4-5.5)** (characterized by an imaginary frequency of -104 cm^{-1}) the amine group of IPP is closer to carbon C5 of the acyl–enzyme complex (C5(EOC)···N7(IPP) distance of 2.72 Å), and the hydrogen atom pointing toward His282 is starting to contract (N3···H7 distance of 1.74 Å versus 2.15 Å in the reactants). In the first intermediate **5.5**, the amine molecule becomes covalently bound to the acyl–enzyme complex forming a tetrahedral intermediate that is negatively charged. The hydrogen from the amine group of IPP still bonds to the nitrogen; this nitrogen atom is thus positively charged. This first step requires an activation energy of 17.3 kcal/mol, and a change of the overall energy of reaction of 17.1 kcal/mol (**Table 5.2**).

The second step of the amidation involves the hydrogen transfer from the nitrogen of IPP toward His282. In the second transition state **TS(5.5-5.6)** (characterized by an imaginary frequency of -387 cm^{-1}) the interaction between the amine group of IPP and the C5 carbon of the acyl–enzyme complex has increased (C5(EOC)···N7(IPP) distance is 1.59 Å compared to 1.76 Å in the previous step), and the hydrogen atom is shared with one of the nitrogen atoms of His282 (N3(His282)···H7(IPP) distance of 1.17 Å versus 1.70 Å in 5.5). The distant from the Nδ atom of His282 to Asp252 is unchanged and thus does not seem to compensate for the acceptance of the additional hydrogen on His282. Following **TS(5.5-5.6)**, the second tetrahedral intermediate is formed (**5.6**, **Figure 5.4a**). In this second intermediate **5.6**, there is a small rotation of the acyl substrate that favors the proximity of the oxygen from Ser155 to the NH group of His282, while it is firmly

attached to the tetrahedral acyl complex (O4(Ser155)–C5(EOC) bond length of 1.51 Å). This final intermediate remains negatively charged, and it is stabilized by two hydrogen bonds provided by the two glycines and the alanine that form the oxyanion hole. This second step requires an activation energy of 6.6 kcal/mol, and a reaction energy of 6.4 kcal/mol, relative to **5.5** (**Table 5.2**).

The third and final step of the amidation mechanism involves the release of the amide product and turnover of the active site. In the third transition state **TS(5.6-5.7)** (characterized by an imaginary frequency of -625 cm^{-1}) Ser155 is partially dissociated from the substrate (O4(Ser155)–C5(EOC) distance of 1.84 Å versus 1.51 Å in **5.6**). The proton that was previously attached to His282 is now shared with Ser155 (H7⋯O4 distance of 1.23 Å and N3(His282)⋯H7(His282) distance of 1.22 Å). In the product of the reaction (**5.7**), Ser155 is finally reprotonated and unbound from the substrate. The substrate is now free to dissociate from the active site, although the carbonyl group remains stabilized by hydrogen bond interactions with the oxyanion hole. This final reaction requires an activation energy of 4.8 kcal/mol and is exergonic by 10.8 kcal/mol, relative to **5.6** (**Table 5.2**).

The complete energy path is presented in **Figure 5.5**. The rate limiting steps of the reaction are **TS(5.2-5.3)** for the acylation step and **TS(5.6-5.7)** for the amidation step; the rate-determining step in the acylation pathway was therefore further investigated via charge modification procedures in order to identify potential residues (hotspots) that could lower this barrier.

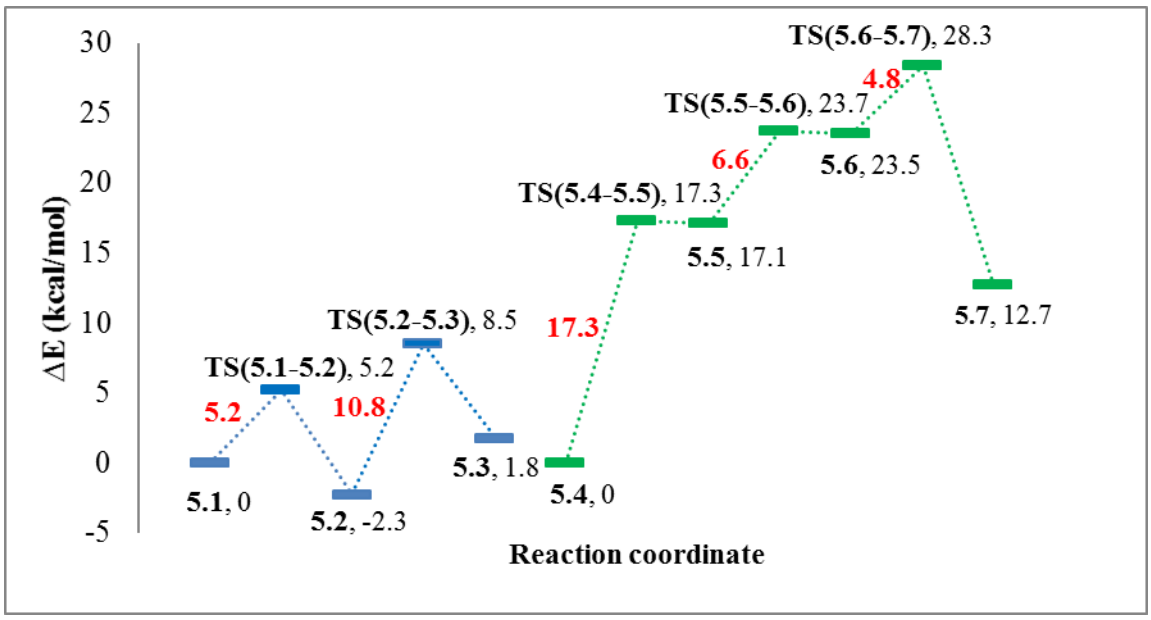


Figure 5.5. Potential energy profile involved in the formation of the acyl–enzyme complex (blue) and the amidation (green). QM/MM electronic energies (ΔE) in kcal/mol are given relative to the reactant state for each step.

5.4 Charge modification procedures results

5.4.1 ElectroScan results on the rate limiting step for the acylation pathway

The impact of a given mutation in the enzyme on the rate limiting barrier between **5.2** and **TS(5.2-5.3)** was estimated by setting to zero the MM charges on the corresponding side chain and recalculation of the energy of the rate limiting barrier corresponding to stationary points. This procedure was repeated for all residues of the enzyme. The results of this 'ElectroScan' for **TS(5.2-5.3)** are presented in **Figure 5.6**. The calculations identified six amino acid position that have a pronounced effect (≥ |1| kcal/mol) on the reaction barrier. These residues are Asp154 (-3.4 kcal/mol), Arg255 (+1.5 kcal/mol), Lys259 (+1.2 kcal/mol), Asp256 (-1.5 kcal/mol), Glu278 (-1.0 kcal/mol) and Glu199 (-1.1 kcal/mol). In these results there is a clear preference for the identification of acidic and basic residues at hotspots, given the 20 potential gene-encoded amino acids. This preference was already observed in previous charge modification studies.[46, 210] These type residues are likely to be ionized at physiological pH, with basic residues positively charged and acidic amino acids negatively charged, it is thus not that surprising that they have the largest impact in term of electrostatics. Furthermore, **TS(5.2-5.3)** is associated with a proton transfer and thus particularly sensitive to a modification of the charge environment. In **Figure 5.6** residues are ordered by their distance to the active site, it is observed that the hotspot residues are not found at a specific distance from the active site, more precisely the hotspot residues can be arranged into three different groups: a first group formed of Asp154 that is located in the active site, a second group formed of

Arg255, Lys259 and Asp256 is located 10-12 Å from the active site and finally a last group (Glu278 and Glu199) is located ~17 Å from the active site.

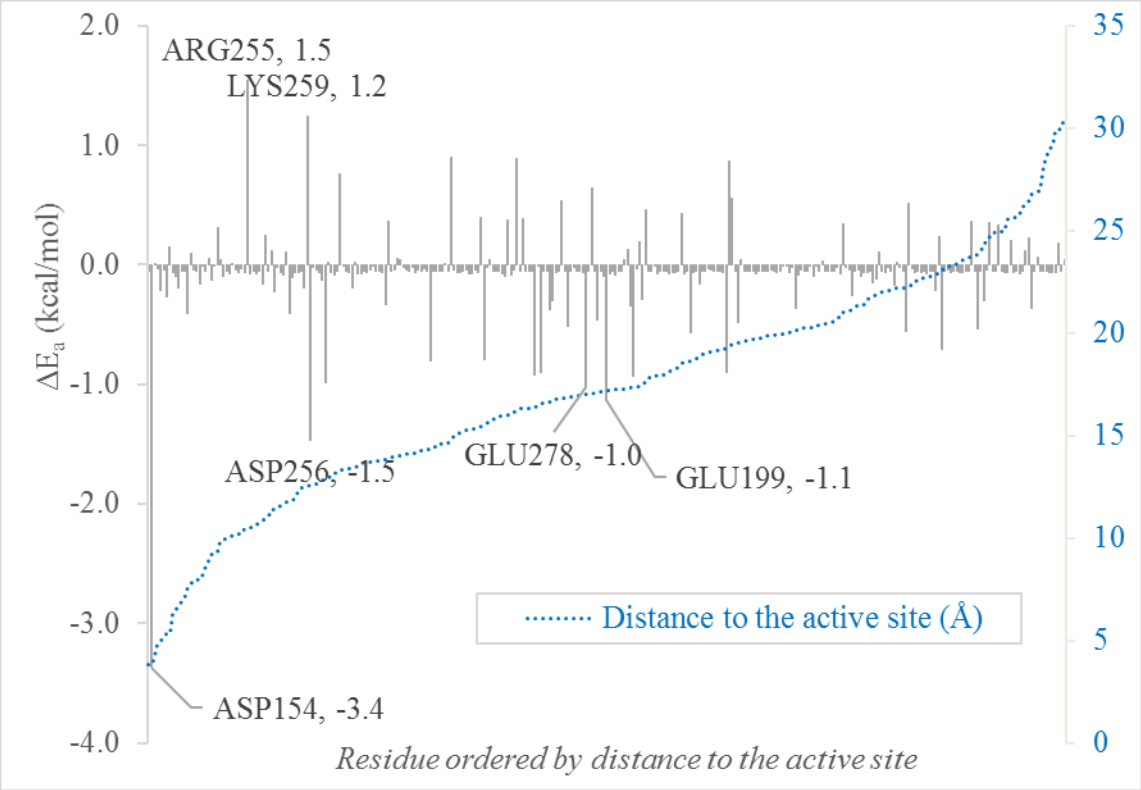


Figure 5.6. 'ElectroScan' results for **TS(5.2-5.3)**.

To have a better understanding of why the charge nullification of these residues have a significant impact on the reaction barrier and why for some the barrier is increased and for other it decreases, we can analyze a 3D representation of the hotspot residues positions compared to the **TS(5.2-5.3)** reaction vector from **Figure 5.7**.

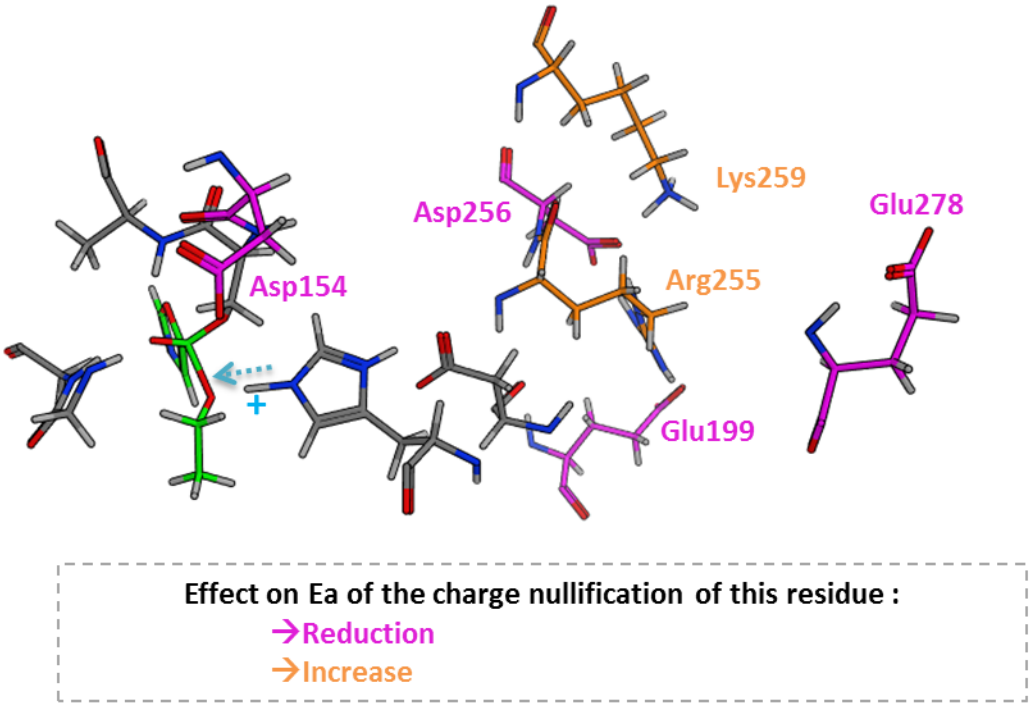


Figure 5.7. Location of the hotspot residues for the **TS(5.2-5.3)** 'ElectroScan'. The proton that is transferred during this step is indicated by a plus sign in cyan, the vector of this transfer is indicated by a cyan dotted arrow. Residues for which charge nullification reduces the energy barrier between **5.2** and **TS(5.2-5.3)** are represented in pink and those for which it increases this barrier are represented in orange.

From Figure **5.7** we can see that residues for which charge nullification has a significant impact on the reaction barrier are located mostly at one side of the reaction vector with the exception of Asp154. More precisely residues for which the charge nullification reduces the energy barrier for this step are located in the opposite direction of the proton transfer vector when they are acidic. The reverse situation can be seen for residues that increase the barrier. The reduction of the energy barrier from the charge nullification

works by reducing the electrostatics energies unfavorable for the proton transfer by suppressing negative charges in the opposite direction of the vector of the proton transfer.

5.4.2 ElectroScan results on other steps

In the present work, **TS(5.2-5.3)** was first chosen for the 'ElectroScan' procedure because it is the rate limiting step of the acylation step of the reaction. However, it should be kept in mind that this is a multistep reaction and that modification that may be favorable for one step may not be for another. To investigate this the 'ElectroScan' procedure was applied on other steps and the results were analyzed. As summarized in **Figure 5.8a**, most of the other steps are also associated with a proton transfer the only exception being **TS(5.4-5.5)**. Interestingly some proton transfers like **TS(5.2-5.3)/TS(5.5-5.6)** and **TS(5.1-5.2)/TS(5.6-5.7)** share the same line but are in opposite directions as indicated by the arrows in **Figure 5.8a-b**. As a result of this observation the effect of the 'ElectroScan' procedure on **TS(5.5-5.6)** compared to **TS(5.2-5.3)** was investigated (**Figure 5.9a**). The effect of 'ElectroScan' on a step that is not associated with a proton transfer, **TS(5.4-5.5)**, was also investigated (**Figure 5.9b**).

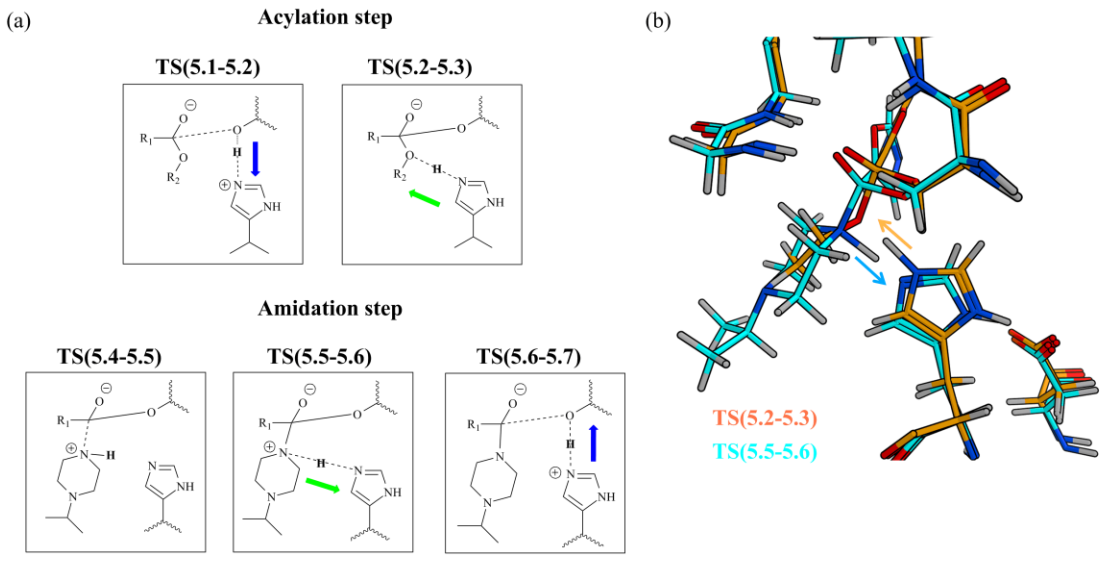


Figure 5.8. Proton transfers during the enzymatic amide bond formation (a) Schematic representation of transition state associated with proton transfer. Proton transfer vectors sharing the same line are represented by the same color and the direction is indicated by an arrow. (b) Superposition of **TS(5.2-5.3)** (orange) and **TS(5.5-5.6)** (cyan). The arrows indicate the proton transfer vector associated with each step.

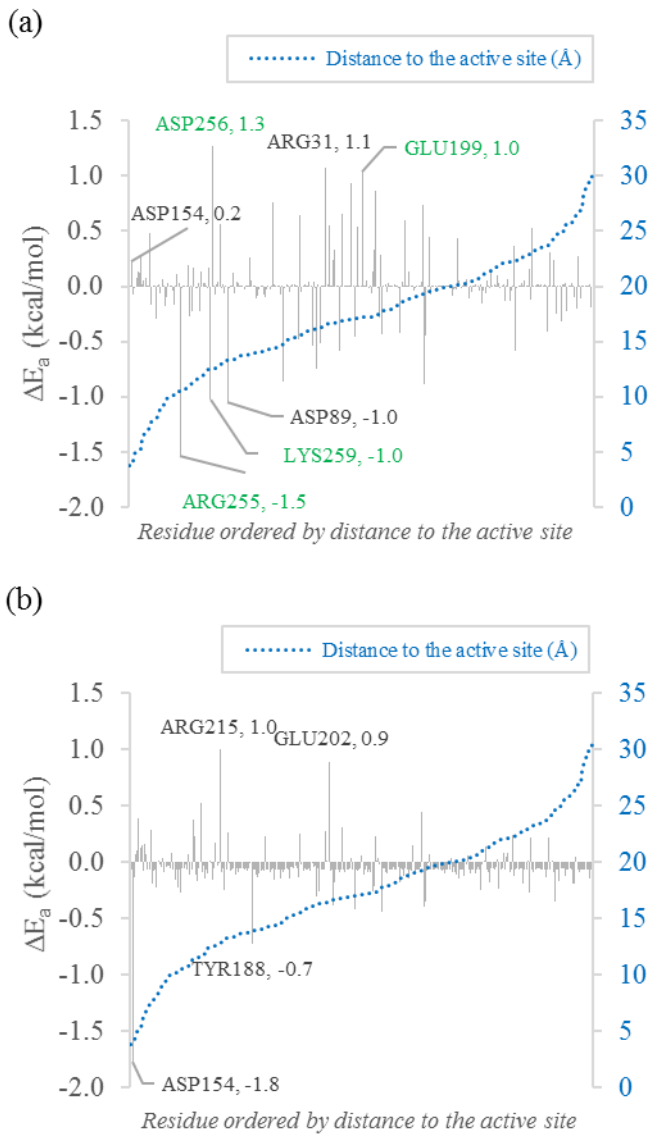


Figure 5.9. 'ElectroScan' results for (a) **TS(5.5-5.6)** and for (b) **TS(5.4-5.5)**. Hotspot residues identified in both **TS(5.2-5.3)** and **TS(5.5-5.6)** steps are represented in green.

The opposite nature of the proton transfer direction of **TS(5.2-5.3)** and **TS(5.5-5.6)** is also found in the 'ElectroScan' results as some hotspot residues are shared between **TS(5.2-5.3)** and **TS(5.5-5.6)** but with the opposite effect; these are Asp256 (-1.5 for **TS(5.2-5.3)**

and 1.3 for **TS(5.5-5.6)**) and Arg255 (1.5 for **TS(5.2-5.3)** and -1.5 for **TS(5.5-5.6)**) for example. But interestingly for Asp154 the charge nullification does not have a particular effect on **TS(5.5-5.6)** (+0.2 kcal/mol) barrier whereas it had a significate effect on **TS(5.2-5.3)** (-3.4 kcal/mol).

The results of the 'ElectroScan' procedure applied to **TS(5.4-5.5)** is presented in **Figures 5.9b**, the calculations identified only one amino acid position that has a pronounced effect ($\geq$ |1| kcal/mol) on the reaction barrier. This residue is once again Asp154. The charge nullification of the partial charges of the side chain of Asp154 reduces the energy barrier between **5.4** and **TS(5.4-5.5)** by 1.8 kcal/mol. Thus, there are less residues for which charge nullification has a significant influence on the barrier than was the case for proton transfer steps **TS(5.2-5.3)** and **TS(5.5-5.6)**.

The opposite nature of the effect of the 'ElectroScan' procedure of the different steps highlights the need for a multistep parameterization strategy for our system. Therefore, rather than identifying hotspots for one specific step another option would be to rank residues according to the sum of their effect on different steps. This approach was taken for steps associated with **TS(5.2-5.3)**, **TS(5.4-5.5)** and **TS(5.5-5.6)** and results for the five residues with the larger magnitude sum are summarized in **Table 5.3**.

Table 5.3. 'ElectroScan' results for **TS(5.2-5.3)**, **TS(5.4-5.5)** and **TS(5.5-5.6)** for residues with the larger absolute magnitude sums.[a]

	ΔE_a			
Residue	**TS(5.2-5.3)**	**TS(5.4-5.5)**	**TS(5.5-5.6)**	**Sum**
Asp154	-3.4	-1.8	0.2	-4.9
Thr186	-0.2	-0.2	-0.3	-0.7
Tyr183	0.2	0.2	0.3	0.6
Arg215	-1.0	1.0	0.6	0.6
Glu202	-0.9	0.9	0.6	0.5

[a] Relative electronic activation energies (ΔE_a) in kcal/mol are given relative to the unchanged electrostatic environment.

In most cases the effect of **TS(5.2-5.3)** and **TS(5.5-5.6)** cancel each other such that the cumulative effect on the barrier is almost always moderate (< |1| kcal/mol) with the exception of Asp154. For Asp154 the cumulative effect of the 'ElectroScan' over **TS(5.2-5.3)**, **TS(5.4-5.5)** and **TS(5.5-5.6)** yields a total effect of -4.9 kcal/mol. This reflects the fact that for this particular residue the charge nullification has a significant stabilizing effect on **TS(5.2-5.3)** (-3.4 kcal/mol) and **TS(5.4-5.5)** (-1.8 kcal/mol) but does not destabilize the proton transfer in the opposite direction (**TS(5.4-5.5)**, 0.2 kcal/mol). The fact that Asp154 is the only residue for which the cumulative effect over three steps has a significant impact on the reaction can be explained by the fact that this residue is located in the active site (**Figure 5.7**).

5.4.3 Asp154 study

Our calculations have thus systematically identified Asp154 as a hotspot. However, it is also informative to study the conservation of this residue in the expression of this enzyme in nature. Thus a BLAST[333] search of the sequence of the carboxylesterase from *Alicyclobacillus acidocaldarius* was performed against Uniprot[334] and the 100 closest sequences were used to perform a Multiple Sequence Alignment (MSA) using the Clustal Omega program.[335] Finally a sequence logo representation for all these MSA were generated using WebLogo 3,[336] the results of which are presented in **Figure 5.10**. From this analysis it can be seen that residue Asp154 is strongly conserved through evolution. It may be conserved because it helps to give the serine the right pK_a properties. If this residue is confirmed to be a good mutant experimentally it will demonstrate the importance of our approach as strongly conserved residues are usually not picked for screening libraries.

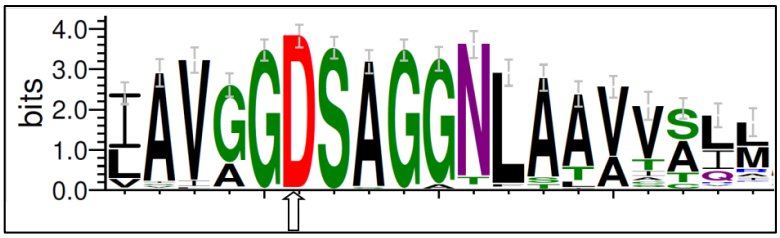


Figure 5.10. WebLogo[336] representation of a MSA of the 100 closest sequences to the sequence of the carboxylesterase from *Alicyclobacillus acidocaldarius* on Uniprot[334] (Asp154 is indicated by an arrow).

Overall the application of the 'ElectroScan' procedure on this enzymatic amide bond formation has allowed us to gain a better understanding of the mechanism by

demonstrating that the arrangement of the amino acids around the binding site helps to support favorable electrostatics for the proton transfer step in the acylation mechanism. However, this study also identified possible limitations of the 'ElectroScan' procedure. This procedure allows for optimizing the electrostatic contributions of the protein only by suppressing unfavorable electrostatics but in the process of optimizing an electrostatic field it should also be possible to introduce new favorable electrostatics interactions. From this finding came the idea of introducing a new procedure: the 'PosiScan'.

5.4.4 PosiScan results

The idea of the 'PosiScan' procedure is to have a modified version of the original 'ElectroScan' procedure where not only is there a charge nullification of the partial charges of the side chains but also an electrostatic 'perturbation' is introduced by assigning a +1 charge at the Cβ that has the property of being a common side chain atom for almost all the amino acids (with the exception of Glycine). The results of this 'PosiScan' procedure applied to the step associated to **TS(5.2-5.3)** is shown in **Figure 5.11** and residues with the larger magnitude $|\Delta E_a|$ results are given in **Table 5.4**.

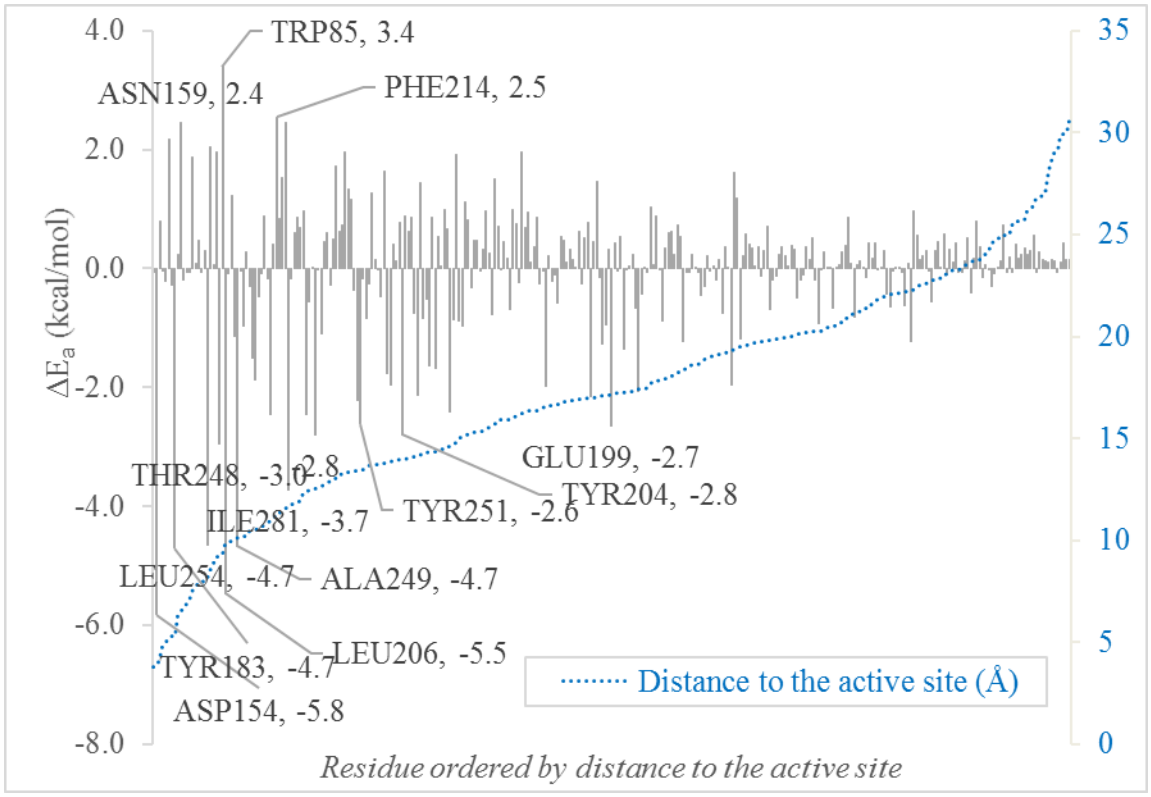


Figure 5.11. 'PosiScan' results for **TS(5.2-5.3)**.

Table 5.4. 'PosiScan' results for **TS(5.2-5.3)** for residues with the larger absolute magnitudes $|\Delta E_a|$.[a]

Residue	**ΔE_a**
Asp154	-5.8
Leu206	-5.5
Tyr183	-4.7
Ala249	-4.7
Leu254	-4.7
Ile281	-3.7
Trp85	3.4
Thr248	-3.0

[a] Relative electronic activation energies (ΔE_a) in kcal/mol are given relative to the unchanged electrostatic environment.

These results demonstrate that the 'PosiScan' approach is able to explore new regions of the protein compared to the 'ElectroScan' by identifying not only acidic/basic residues but also neutral residues, such as Leu206 (-5.5 kcal/mol) and Tyr183 (-4.7 kcal/mol). It appears that there is a stronger distance dependence effect on the intensity of the peaks (**Figure 5.11**) than there is in the 'ElectroScan' results (**Figure 5.6**). It could be hypothesized that as there are more peaks of strong intensity, the Coulombic type distance dependence of the electrostatic interaction becomes more evident. However this distance dependence observation results from one enzyme and can thus not be generalized until tested on other systems.

The highest hotspot is once again Asp154 (-5.8 kcal/mol). Strongly negative ΔE_a results suggest that the introduction of a positive charge at this location (by mutating to lysine or arginine for example) could potentially lead to a gain of activity. We also expect that having strongly positive ΔE_a is a sign that the introduction of a negative charge at this location (by mutating to aspartate or glutamate for example) could also lead to a gain of activity. We have tested this hypothesis on Trp85 for which the 'PosiScan' approach gives a ΔE_a of 3.4 kcal/mol (**Table 5.4**). The *in silico* mutation of Trp85 to an aspartate gives a ΔE_a of -3.3 kcal/mol and thus lends support to this hypothesis.

The 'PosiScan' approach is meant to be a fast screening method that could then be followed up with a real computational mutation. This would be the next step in evaluating the feasibility of such a mutation. It should be checked that the identified hotpot residue is not implicated in stabilizing interactions that could be disrupted by a mutation. Also, there is a considerable variation in the sizes of amino acids and thus some mutations can be easier than others. It should thus be aimed to replace residues of similar sizes, one good example being the aspartate to asparagine or the glutamate to glutamine mutation as these occupy almost the same space. The molecular weight (MW) of an amino acid is roughly proportional to its size and can thus be used to compare them (**Figure 5.12**).

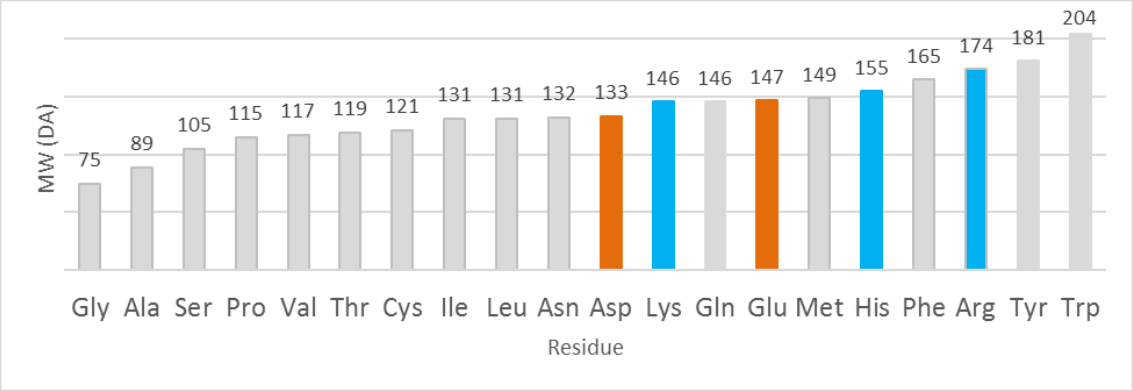


Figure 5.12. Molecular weight (MW) of amino acids. Basic residues are colored in blue and acidic in orange.

From **Figure 5.12** we can see that basic and acidic residues are part of the right hand half of residues having the highest molecular weight; a mutation to one of these residues will probably not always be possible. It can also be observed that lysine has a lower MW than arginine (146 DA and 174 DA, respectively), a mutation to lysine should thus be generally privileged over arginine. The same can be observed between aspartate (133 DA) and glutamate (147 DA). Histidine is also a basic residue but with a side chain pK_a of approximately 6.0 it is not certain that it will be charged under physiologically relevant pH values. Among the basic/acidic residues aspartate has the lower MW and could thus be the easiest to accommodate. This could suggest that a 'NegaScan' approach (with a -1 charge on Cβ) could be prioritized over a 'PosiScan'.

5.5 Conclusion

The goal of the study was to elaborate and test a protocol for the identification of mutations that could enhance an enzyme activity. For this, an amide bond formation reaction catalyzed by a carboxylesterase was used as case study. The underlying mechanism of the reaction was first investigated using QM/MM. A catalytic mechanism was suggested with a two-step acylation followed by a three step amidation.

The so-called ElectroScan technique was then applied to the rate-limiting step of the acylation step. Partial charges of single residues side chains were set to zero and the activation energy re-evaluated in the new environments. The results gave 6 hotspots (with $\Delta E_a \geq |1|$ kcal/mol) in Asp154, Arg255, Lys259, Asp256, Glu278 and Glu199. The subsequent application of the 'ElectroScan' technique to other steps of the reaction (first and second transition states of the amidation step) revealed that the applicability of this methodology for this system was found to be limited by the reaction mechanism. More precisely, **TS(5.2-5.3)** and **TS(5.5-5.6)** are associated with proton transfer reactions that share the same line but in opposite direction, thus depending on the direction of the proton transfer different amino acids are identified and side chains charge nullification effects are mutually cancelled. However, because of it position in the active site a hotspot residue was identified in Asp154.

Finally, the 'PosiScan' approach was also tested on the same reaction. It was found that this approach could be a possible strategy to introduce new favorable electrostatic

interactions in a protein but that a ‘NegaScan’ approach could be prioritized over the ‘PosiScan’ due to steric reasons.

Overall, the second initial aim of creating and testing a QM/MM protocol for hotspot identification was achieved in this chapter. The next chapter explores how this protocol could be further speed-up.

6. Computational investigation of a lipase catalyzed amide bond formation reaction

6.1 Introduction

An Enzyme Panel is a set of diverse enzymes (*e.g.*, phylogenetically diverse imine reductases) that can perform a specific transformation (*e.g.*, synthesis of amines) on a pharmaceutically relevant molecular space.[98] Enzyme panels are created to be ready to use toolkits for organic synthesis. They are typically prepared as plates that can then be used for screening on a desired transformation – in a similar manner to how a traditional chemical catalyst screening is carried out. GlaxoSmithKline (GSK) has been working on the in-house creation of enzyme panels for high priority transformations, examples currently addressed are shown in **Figure 6.1**.[98] The in-house creation of these panels allows freedom-to operate status and avoids potentially complex and expensive IP agreements.[98]

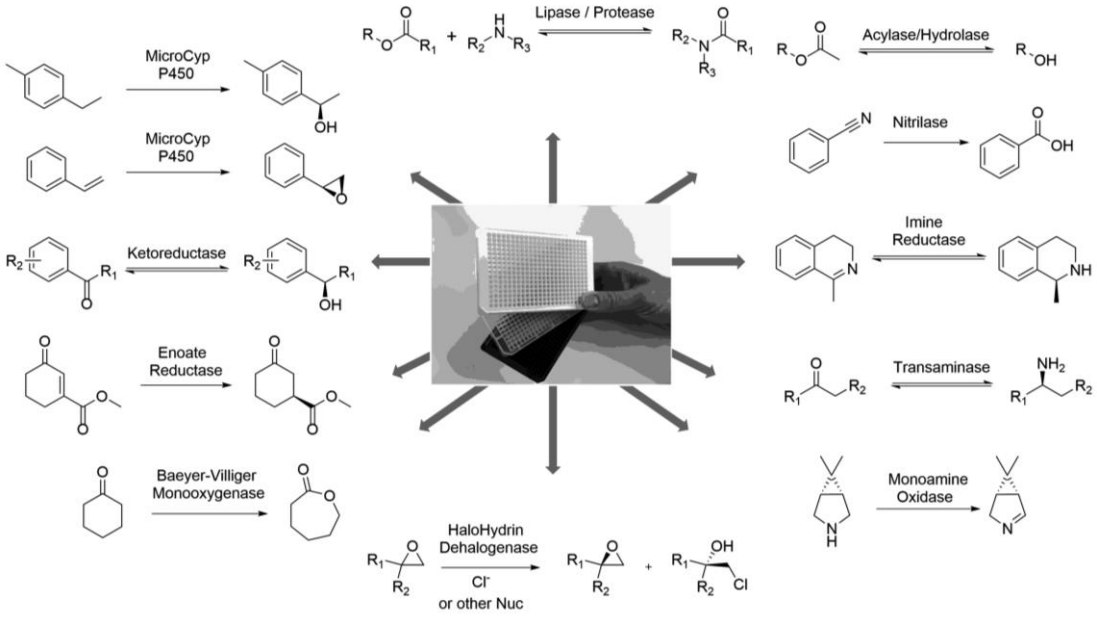
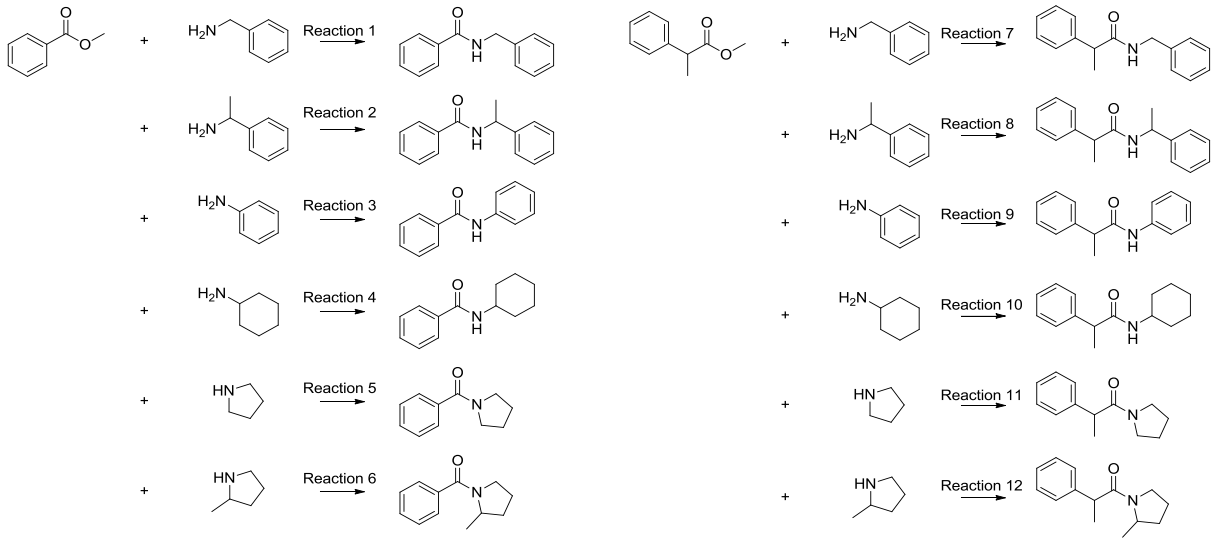


Figure 6.1. Examples of enzymatic transformations currently addressed by customised panels at GSK. Figure taken with permission from [98].

Among those high priority transformations there is the amide bond formation (ester aminolysis) by a lipase enzyme panel. To address this top priority transformation, GSK and Pfizer have decided to join forces through the collaborative creation of an enzymatic amidation platform. A panel of 12 reactions described in **Scheme 6.1** was chosen as the target for this project. The activity of different lipases was screened on this panel and it was a lipase (EC 3.1.1.3) from *Pseudomonas batumici* (PBL) that gave the most interesting results (K. Brown, GSK, personal communication). The initial activities of PBL on the reaction panel described in **Scheme 6.1** were communicated by our experimental collaborators and are summarized in **Table 6.1**. The results show some initial activity for reaction 1 to 6 and very little for reaction 7 to 12. To improve the activity of PBL on the panel reaction a directed evolution strategy was conducted.[337]



Scheme 6.1. Reaction panel

Table 6.1. Ester aminolysis with wild type PBL.[a, b]

Reaction	Ester	Amine	%conversion
1	methyl benzoate	benzylamine	38.09
2	methyl benzoate	(*rac*)-1-phenylethanamine	6.03
3	methyl benzoate	aniline	8.12
4	methyl benzoate	cyclohexylamine	29.07
5	methyl benzoate	pyrrolidine	21.68
6	methyl benzoate	(*rac*)-2-methylpyrrolidine	2.64
7	methyl 2-phenylpropanoate	benzylamine	0.07
8	methyl 2-phenylpropanoate	(*rac*)-1-phenylethanamine	0.00
9	methyl 2-phenylpropanoate	aniline	0.75
10	methyl 2-phenylpropanoate	cyclohexylamine	0.07
11	methyl 2-phenylpropanoate	pyrrolidine	0.02
12	methyl 2-phenylpropanoate	(*rac*)-2-methylpyrrolidine	0.00

[a] Data obtained from K. Brown, GSK, personal communication

[b] %conversion means percentage of substrate transformed after 24h

The initial step of a directed evolution process is the creation of a library of variants. In this chapter, the hotspot identification QM/MM protocol established in **Chapter 5** is applied to reaction 1 (**Scheme 6.1**) catalyzed by PBL to aid the creation of such library. As PBL does not have a crystal structure a homology model strategy is first conducted. This chapter also describes the effect of deleting the solvent from the QM/MM model on both PBL mechanistic studies and hotspot investigation output; this is done by comparing to the results obtained in water and toluene.

6.2 Computational methods

6.2.1 Homology model

As no crystal structure was available for the enzyme a model was created. To identify a template structure, the PBL sequence was used to run a BLAST[333] search against the Protein Data Bank (PDB), the top five results are presented in **Table 6.2**.

Table 6.2. PDB BLAST results using sequence from lipase of *Pseudomonas batumici* (accessed on 01/12/2017).

PDB code	Organism	Score	E-value
3W9U	*Proteus mirabilis*	603	2.20983E-62
4GW3	*Proteus mirabilis*	603	2.42225E-62
4GXN	*Proteus mirabilis*	603	2.42225E-62
4HS9	*Proteus mirabilis*	588	1.07889E-60
1EX9	*Pseudomonas aeruginosa*	541	3.77706E-55

The first four PDB structures with the higher score were associated with the same lipase from *Proteus mirabilis* (PDB ID : 3W9U[338], 4GW3[339], 4GXN[339] and 4HS9[340]) and the fifth hit with a lipase from *Pseudomonas aeruginosa* (PDB ID : 1EX9[341]) (**Table 6.2**); all structures are part of the I.1 bacterial lipase family. The scores and E-values for PDB structures of lipase from *Proteus mirabilis* were 2.20E-62/603 (3W9U), 2.42E-62/603

(4GW3), 2.42E-62/603 (4GXN) and 1.07E-60/588 (4HS9), respectively. The Expect value (E) is a parameter that describes the chance to obtain one match with a similar score simply by chance. The top three PDB structures (3W9U, 4GW3 and 4GXN) have thus the same score and a slightly smaller E-value for 3W9U which is due to a slightly smaller sequence (a polyhistidine-tag is absent in this protein and present in the others). The last PDB structure (4HS9) corresponds to a methanol tolerant mutant (13 mutations were introduced) of the *Proteus mirabilis* lipase which explains it lower score.[340] The E-value and score for PDB structure of lipase from *Pseudomonas aeruginosa* were of 3.77E-55/541 (1EX9) which gives a 62 point score difference with the higher scoring *Proteus mirabilis* PDB structures.

The lipase from *Proteus mirabilis* is in a closed conformation (two typical α helices of the I.1 bacterial lipase family are interacting together and close the entrance to the active site) in the four PDB structures (3W9U, 4GW3, 4GXN and 4HS9) whereas the lipase from *Pseudomonas aeruginosa* in an open conformation (1EX9) (**Figure 6.2**). It is believed that lipases have a closed conformation and open up when they are at the interface between the aqueous and the oil phase – "interfacial activation".[342] However, in *Proteus miribalis* lipase the ligand was found bonded in the "closed" conformation typical of this family but in this structure there is a wide active site that is solvent accessible (**Figure 6.2b**).[339] The lipase from *Proteus miribalis* thus seems to represent a particular case of lipase without interfacial activation, we thus decided to focus the rest of our study on the homology model from *Pseudomonas aeruginosa* lipase. Homology modelling of PBL was conducted in MOE[268] (Molecular Operating Environment) using the crystal structure of

Pseudomonas aeruginosa lipase (PAL, PDB ID: 1EX9; Resolution: 2.54 Å) which has a 44.6% sequence identity with PBL. The PAL structure has a calcium ion that stabilizes the loop containing the catalytic residue His251. This calcium ion is coordinated by two aspartic acid residues (Asp209 and Asp253). A calcium ion was thus placed at equivalent position in the homology model which corresponds to residues Asp220 and Asp265.

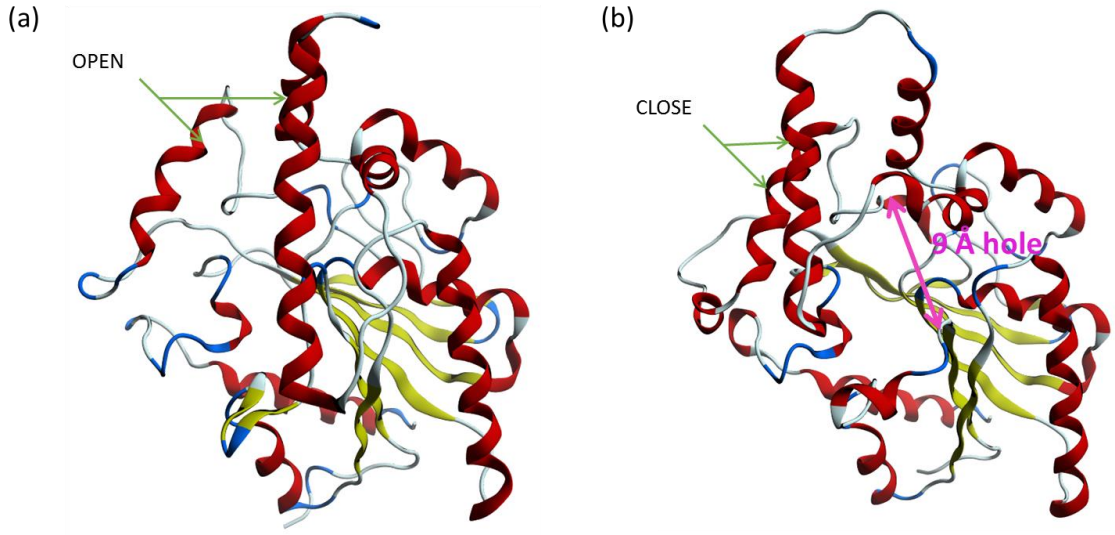


Figure 6.2. View of the α-carbon backbone trace (schematic diagram) of the inhibited structures of lipase from (a) *Pseudomonas aeruginosa* (PDB ID: 1EX9) and from (b) *Proteus mirabilis* (PDB ID: 4GXN). The "close" or "open" conformation of members of the I.1 bacterial lipase family is usually defined by the positioning of two α helices : when they interact together they typically close the entrance of the active site which does not allow any ligand to enter the active site and is thus called a"close" conformation. In *Proteus mirabilis* lipase, even though there is a typical "close" conformation arrangement there is also a 9 Å hole which can allow a ligand to enter the active site even in this conformation.

6.2.2 Systems preparation

The first ligand of the reaction, methyl benzoate (MBZ), was manually added to the model. The carbonyl of MBZ was positioned in a configuration that would enable a nucleophilic attack of Ser83. The parameters for MBZ were taken from the OPLS3[331] force field. In order to obtain a good starting point for the transition state search, a conformational search with constraints on the distances between the carbonyl carbon of MBZ and the oxygen of Ser83 was run. From this search two conformations were obtained but only one was a good starting point for the acylation reaction and was thus used for the rest of the calculations.

The system was solvated as described in Section 3.3.4 and neutralized by random addition of 8 chloride atoms. The system was then gradually relaxed using a standard protocol implemented in Desmond.[270-271] A molecular dynamics (MD) simulation was performed for 25 ns using Desmond[270-271] in order to further relax the system and to assess the stability of the homology model.

Because the reaction takes place in toluene the stability of the homology model in this solvent was also evaluated. However, as toluene solvation was not available in the system builder of Desmond[270-271] a custom solvent model was created. A single molecule of toluene was first parametrized in Ligprep[343] using the OPLS3[331] force field. Atom-centered partial charges were then derived by fitting the electrostatic potential using *ab initio* HF/6-31+G* level with Jaguar[344] in the gas phase. This molecule was then duplicated several times in order to build a periodic box unit with 125 molecules and a

cubic shape. The box was equilibrated for 5 ns by a molecular dynamics simulation within the NPT ensemble, with the same ensemble method and cut-off treatment as described in Section 3.3.4. No specific rules were used to compute interactions with and between solvent molecules. The final system had a density of 0.87 g/mL in good agreement with experimental value of 0.87 g/mL.[345]

Another system was built by solvating the homology model in a orthorhombic box of toluene with a 10Å buffer between the solute and the box boundary in each direction, using the parametrization described above. This system was also neutralized by random addition of 8 chloride atoms. It should be noted that chloride atoms are insoluble in toluene but that molecular dynamics simulations, particularly when the PME method is employed, need a neutral system and this is why they were added to the system. The relaxation of the "toluene" system required a more careful protocol than for water – when a toluene molecule is removed during the system set up it can leave a larger vacuum that can ultimately cause the failure of a simulation due to system collapse. To avoid this happening, a short (400 ps) NVT simulation with a standard Desmond pre-relaxation protocol was run to homogenize the system. Toluene molecules come closer to the protein and leave some empty spaces in the box (**Figure 6.3**). The following step needed is a 400 ps NPT simulation. During this step the volume of the box decreases and a homogeneous system is obtained. It is worth noting that the protein is frozen during this procedure. The system is then ready for a production run; in this study a 25 ns NPT simulation was ran. At the end of each of these three MD simulations different properties (potential energy,

pressure, temperature and volume) were calculated and analyzed to confirm that the MD simulation was at equilibrium (Appendix Figure A.3.1).

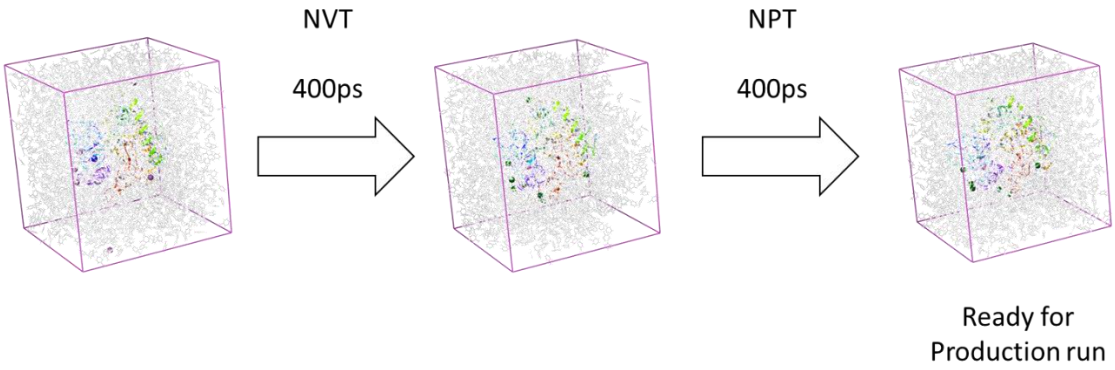


Figure 6.3. Equilibration protocol for toluene model.

6.2.3 QM/MM Methodology

For each model (water and toluene) one snapshot was extracted from the MD output and prepared for QM/MM calculations. A third model (the "vacuum model") was prepared by starting with the water model snapshot and deleting all the water molecules. This model was investigated as an option to accelerate the hotspot identification. The selected snapshot was then MM minimized by using the truncated Newton method[265] implemented in Impact[306]. This post-equilibration, minimized structure, which represents the enzyme-substrate (ES) complex, was used as the starting structure for the QM/MM calculations.

The QM region for the study includes all hypothetical reacting species (Ligand, Ser83 and His263) and stabilizing residues (Met17, Gln84 and Asp241). This QM region is

represented in **Figure 6.4**; it contains 77 atoms in the acylation step and 88 in the amidation step due to the difference of ligand present in the active site.

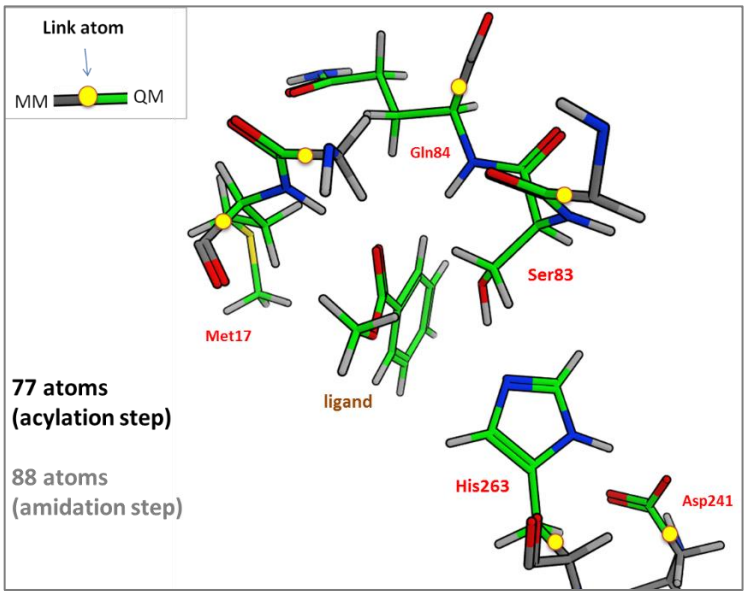


Figure 6.4. QM/MM partitioning use for the calculations (acylation step).

All atoms beyond 15 Å from the reactant were kept constrained during the QM/MM simulations in order to speed up the calculations. The equilibrated ES complex was optimized with QM/MM calculations. The potential energy surface (PES) for the reaction was explored starting from this optimized structure of the reactant as described in the method chapter Section 3.3.6.

To evaluate the electrostatic impact of each residue of the protein on the reaction barrier the 'ElectroScan' procedure as described in Section 3.3.7 of the method chapter was applied. For the first time the 'NegaScan' procedure was also applied.

6.3 Homology model selection

A homology model of PBL was constructed based on the crystal structures of *Pseudomonas aeruginosa* lipase (PAL, PDB ID: 1EX9). The stability of this model was evaluated through a 25 ns molecular dynamics simulation in water. The structure of the model before and after the MD can be found in **Figure 6.5a** and the secondary structure analysis through the simulation in **Figure 6.6a**.

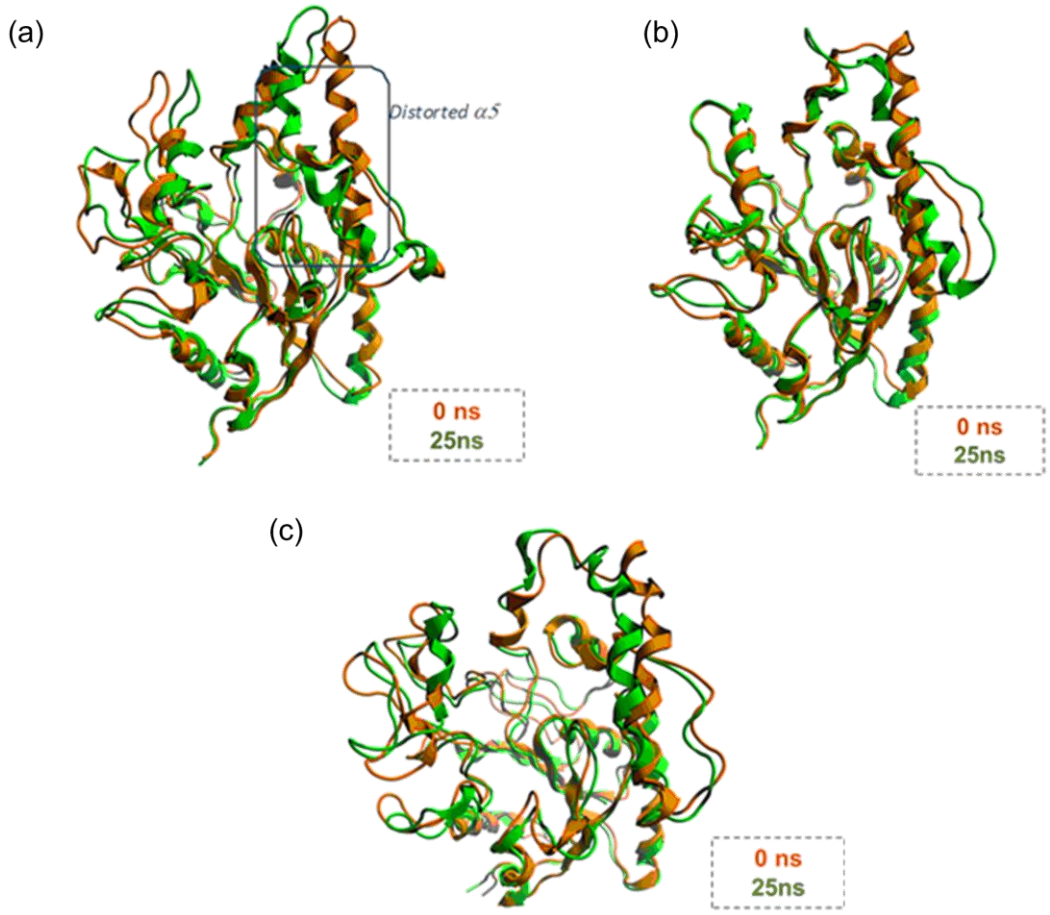


Figure 6.5. Results of the molecular dynamics stability experiment for (a) homology model 1 in water, (b) homology model 2 in water, (c) homology model 2 in toluene. Orange 0ns, green 25 ns.

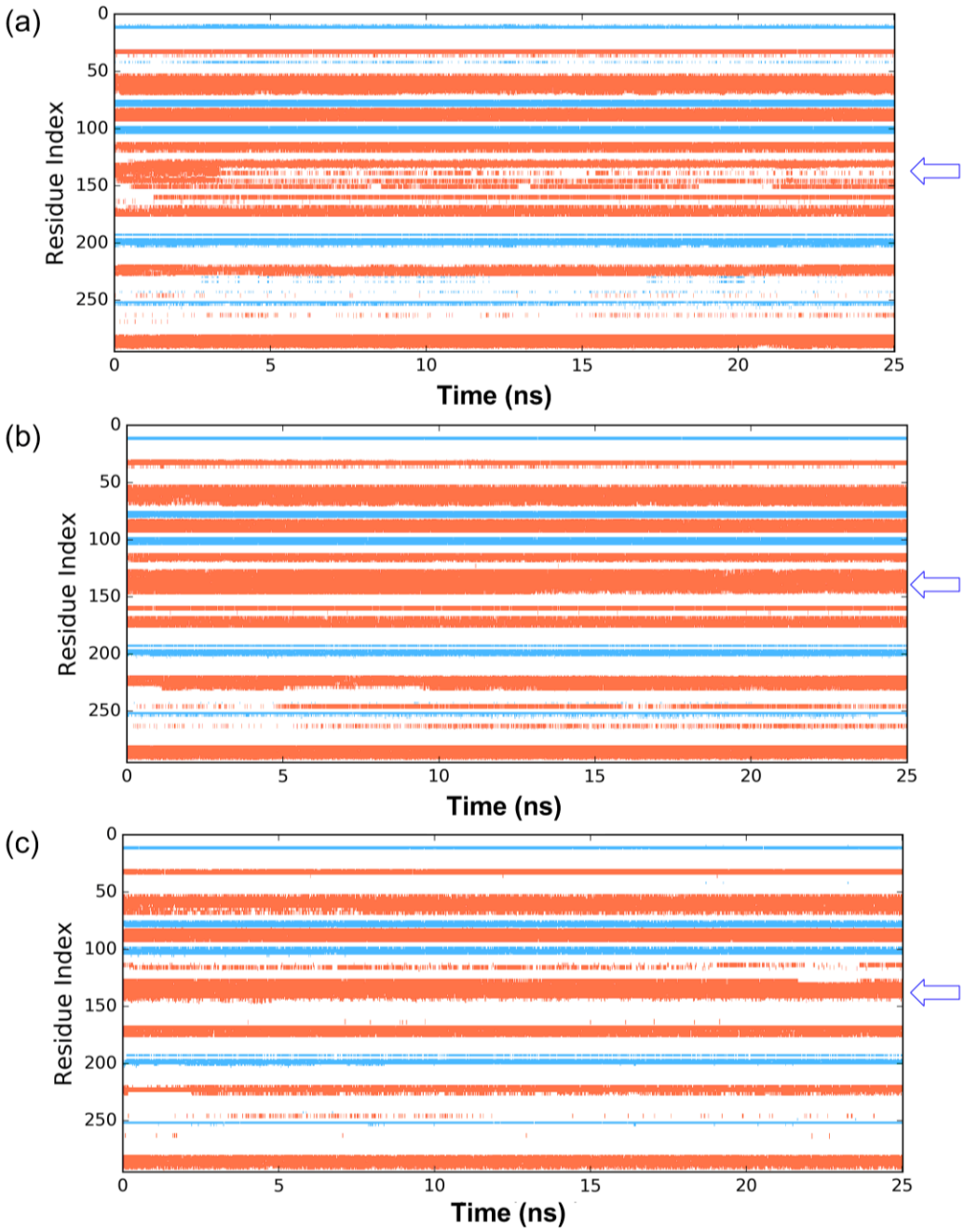


Figure 6.6. Secondary structure analysis for (a) homology model 1 in water, (b) homology model 2 in water, (c) homology model 2 in toluene. Helix 5 is indicated by a blue arrow. β-stands are represented in blue and α-helices in orange.

It can be observed from these results that one of the α-helices (α5) is distorted during the 25 ns simulation. This secondary structure is located between residue 128 and 149 and from **Figure 6.6a** we can see that the α-helix shape is lost after 4 ns. Without knowing the native structure of PBL it is difficult to know if this α-helix unfolding is the consequence of a good refinement by the MD or a poor starting homology model structure.

Considering these results, a different strategy to build the homology model was explored. First a BLAST[333] search of the sequence of PBL was performed against Uniprot[334] and the 250 closest sequences were used to perform a Multiple Sequence Alignment (MSA) using the MUSCLE algorithm.[346] This gave a different alignment between PBL and PAL which was used to build a new homology model (HM2). The stability of HM2 was also evaluated through molecular dynamics simulation and results are presented in **Figure 6.5b** and **Figure 6.6b**. These results show that HM2 retains all its secondary structures during the entire length of the simulation. Because of this stability of HM2 in water more confidence was given to this model than to the first homology model and thus it was selected for the rest of the studies. However, as the reaction takes place in toluene the stability of HM2 was also evaluated in this solvent and results are presented in **Figure 6.5c** and **6.6c**. HM2 also retains all its secondary structures during the 25 ns molecular dynamics simulation in toluene solvent. The stability of these models is further confirmed by backbone RMSD traces shown in **Figure 6.7**.

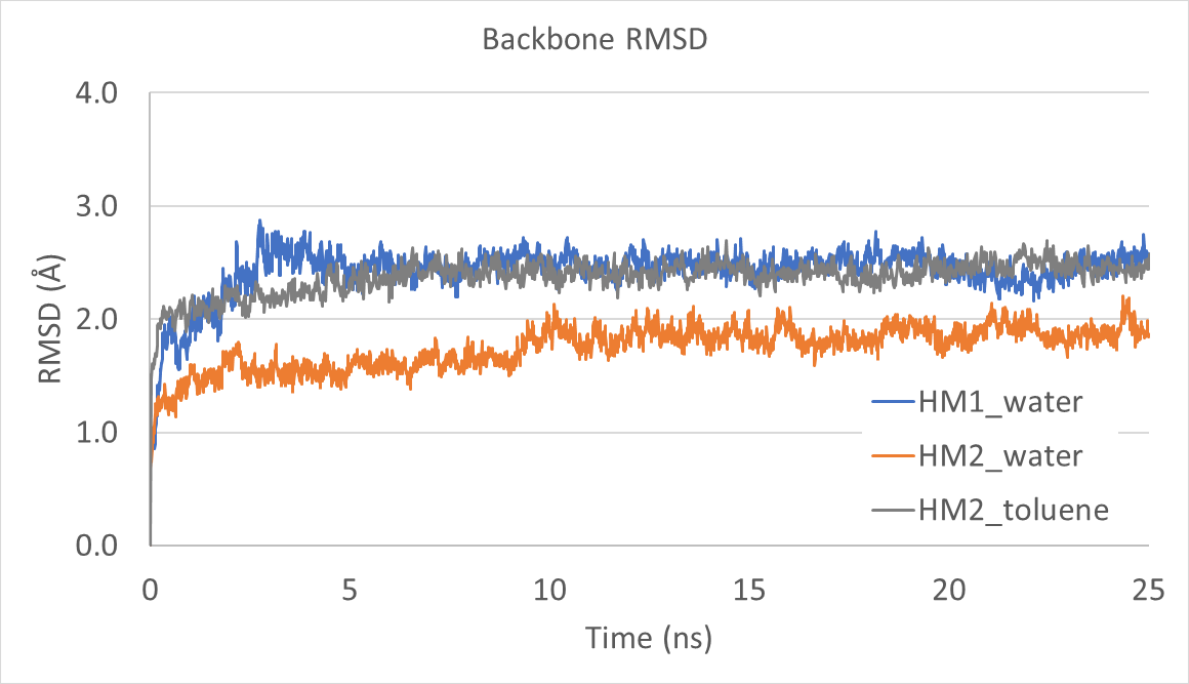


Figure 6.7. Coordinates RMSD of MD-simulated for homology model 1 in water, homology model 2 in water and homology model 2 in toluene. The RMSD is calculated on the protein backbone.

6.4 Proposed catalytic mechanism for the acylation step of the lipase catalyzed amide bond formation reaction

In this work, we unravel the catalytic mechanism of an amide bond formation catalyzed by a lipase. The reaction follows a two-step mechanism: first acylation and then amidation through a tetrahedral intermediate.

In order to clearly differentiate the intermediates from the different models, a labeling scheme is introduced where the abbreviated name of the model (t for toluene, w for water and v for vacuum) precedes the number of the intermediate. Thus, we have for example **t6.1** that stands for intermediate **6.1** of the toluene model. If no letter precedes the intermediate number it references to the intermediate in three cases.

6.4.1 Formation of the Acyl–Enzyme Complex in toluene

The results for the formation of the acyl–enzyme complex requires a two-step mechanism: first the nucleophilic attack of Ser83 to the substrate and second the concomitant formation of the first product of the reaction, an alcohol molecule (**Figure 6.8a**). A schematic representation of the two transition states of the reaction (t**TS(6.1-6.2)** and t**TS(6.2-6.3)**) including only the closest atoms around substrate, is given in **Figure 6.8b-c**. **Table 6.3** shows the relatives energies and the most important distances involved in the formation of the acyl–enzyme complex.

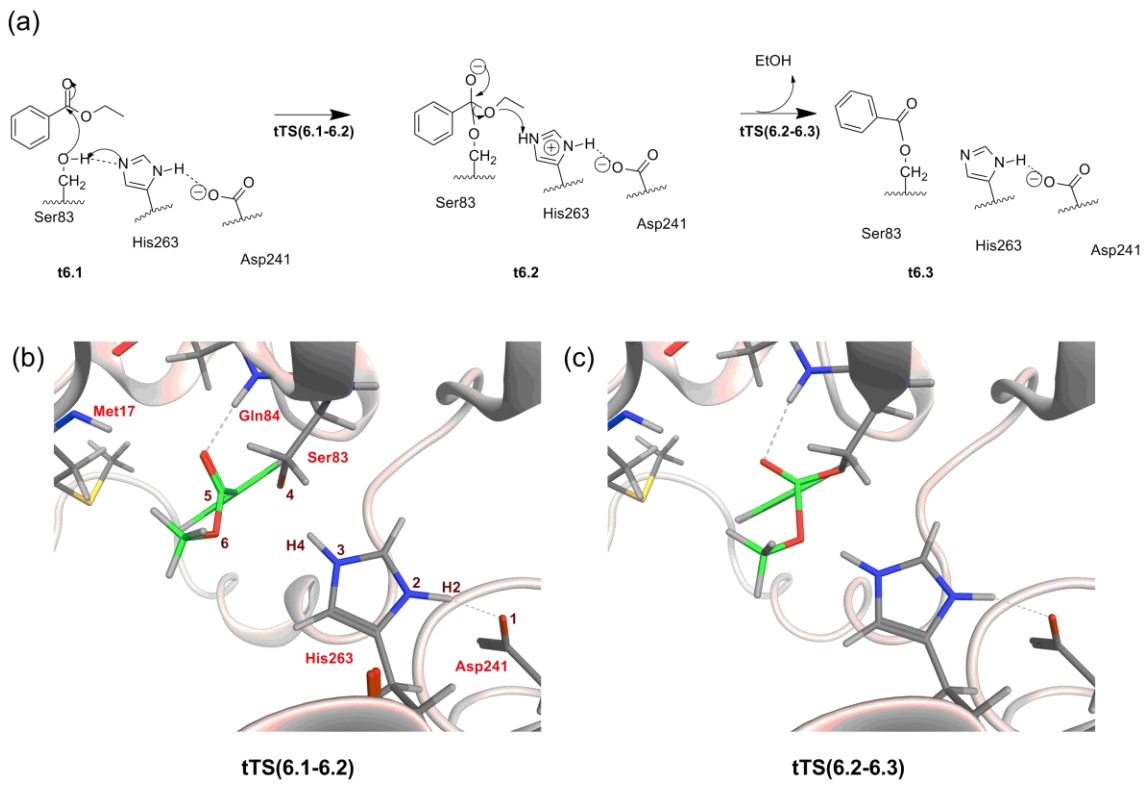


Figure 6.8. Schematic representation of the acylation pathway in toluene and the corresponding transition states structures.

Table 6.3. Relative energies and key distances during the formation of the acyl–enzyme complex in toluene.[a, b]

	ΔE (kcal/mol)	distances (Å)						
		O1···H2	**H2–N2**	**N3···H4**	**H4–O4**	**H4···O6**	**O4···C5**	**C5–O6**
t6.1	0	1.66	1.05	1.8	1	3.7	3.02	1.34
tTS(6.1-6.2)	39.1	1.51	1.1	1.12	1.71	2.77	1.84	1.38
t6.2	27.7	1.51	1.1	1.03	1.8	2.82	1.59	1.44
tTS(6.2-6.3)	41.0	1.68	1.07	1.18	2.54	1.38	1.43	1.77
t6.3	12.7	1.74	1.05	1.93	2.7	0.98	1.35	2.67

[a] Electronic energies (ΔE) in kcal/mol are given relative to the energy of **t6.1**.

[b] Results obtained using B3LYP/6-31G*/OPLS2005 as level of theory

In the starting enzyme-substrate complex (**t6.1**), the carbonyl group establishes two hydrogen bond interactions with the backbone nitrogen atoms of Met17 and Gln84 of the oxyanion hole region (2.47 and 2.04 Å respectively). Ser83 is close to carbon C5 of the substrate (O4(Ser83)···C5(MBZ) distance of 3.02 Å, atom numbering given in **Figure 6.8b**) and establishes a hydrogen bond with His263 (H4(Ser83)···N3(His263) H-bond length of 1.8 Å). His263 also establishes another hydrogen bond with Asp241 (O1(Asp241)···H2(His263) H-bond length 1.66 Å). In the first transition state **tTS(6.1-6.2)** (characterized by an imaginary frequency of -149 cm^{-1}) the proton that was bonded to Ser83 is almost completely transferred to His263 (O4(Ser83)···H4(Ser83) = 1.71 Å and N3(His263)–H4(Ser83) = 1.12 Å). His263 is thus protonated and positively charged which is compensated by a shorter distance between Asp241 and His263

(O1(Asp241)···H2(His263) is 1.51 Å in the **tTS(6.1-6.2)**, compared to 1.66 Å in the reactant complex **t6.1**). All of this gives a stronger nucleophilic character to Ser83, which is closer to the C5 carbon of MBZ (O4(Ser83)···C5(MBZ) = 1.84 Å versus 3.02 Å in the reactant complex). In the first intermediate (**t6.2**), the proton that was previously bonded to Ser83 is now attached to His263, and Ser83 becomes covalently attached to the substrate at carbon C5 (O4(Ser83)–C5(MBZ) bond length of 1.59 Å). This first intermediate of the reaction corresponds to a tetrahedral intermediate which is negatively charged but stabilized in the active site by the hydrogen bonds that are provided by the oxyanion pocket formed by residues Met17 and Gln84 backbone. This first step requires an activation energy of 39.1 kcal/mol, and is endothermic with a reaction energy of 27.7 kcal/mol (**Table 6.3**).

To obtain the acyl–enzyme complex a second step is required. In the second transition state **tTS(6.2-6.3)** (characterized by an imaginary frequency of -525 cm^{-1}) the oxygen atom O6 begins to dissociate from the tetrahedral intermediate (O6(MBZ)–C5(MBZ) bond length of 1.77 versus 1.34 Å in the reactant **t6.1**) and the ε proton of His263 is being transferred to O6 (N3(His263)–H4(His263) = 1.18 Å and O6(MBZ)···H4(His263) = 1.38 Å). This proton transfer is associated with a weaker interaction between His263 and Asp241 (O1(Asp241)···H2(His263) 1.68 Å versus 1.51 Å in the intermediate). In the product of this reaction **t6.3** the acyl–enzyme is formed and the first product of the reaction, an alcohol molecule is released. In the acyl complex the distance between C5 and the oxygen atom of Ser83 is shorter (O4(Ser83)–C5(MBZ) bond length of 1.35 Å) and the hydrogen bond between the other oxygen atom of the substrate and the two

backbone nitrogen atoms of the residues that form the oxyanion hole (Met17 and Gln84) remains present. This step requires an activation energy of 13.3 kcal/mol relative to the previous intermediate **t6.2**, and a change of the overall energy of reaction of -15.0 kcal/mol.

Given the endothermicity of the first step in this reaction, the overall barrier for the acylation reaction is 41.0 kcal/mol. Furthermore, the product of this reaction is formed in an overall endothermic reaction of 12.7 kcal/mol, indicating that the acyl product is very short-lived (**Figure 6.9**).

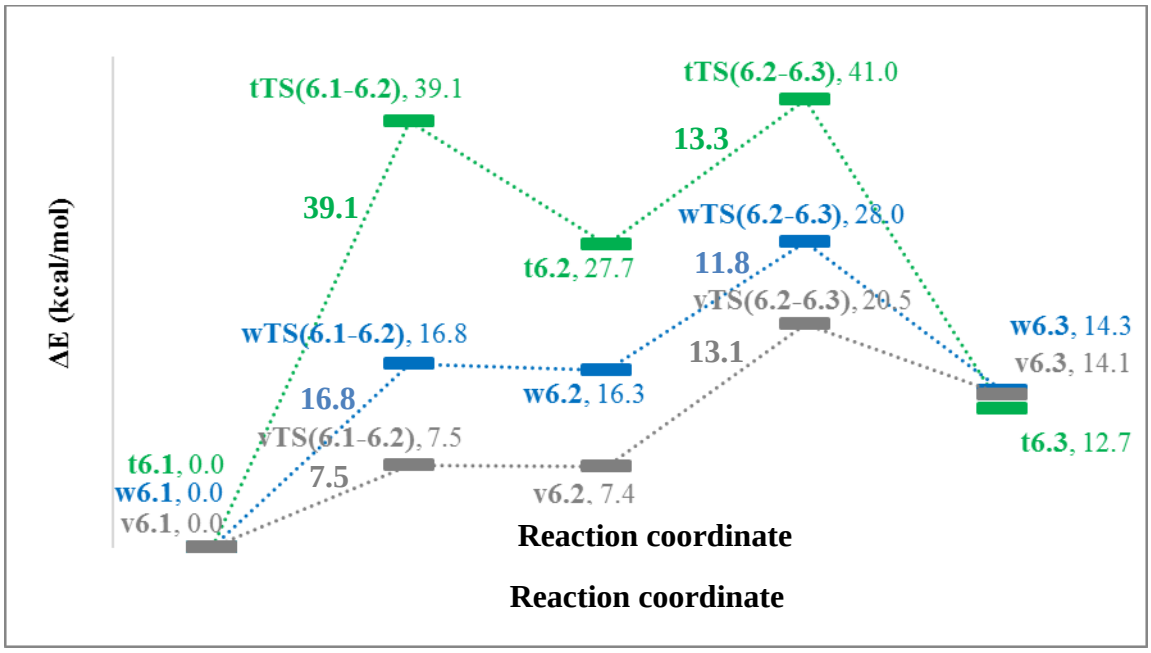


Figure 6.9. Potential energy profile involved in the formation of the acyl–enzyme complex in toluene (green), in water (blue) and in vacuum (grey). QM/MM electronic energies (ΔE) in kcal/mol are given relative to the reactant state for each step.

6.4.2 Formation of the Acyl–Enzyme Complex in water and in vacuum

The results for the formation of the acyl–enzyme complex obtained from water and vacuum models also follows the same reaction mechanism as that presented for toluene, **Figure 6.8a**. **Tables 6.4** and **6.5** show the variation, relative to the toluene model (**Table 6.3**), in the most important distances involved in the formation of the acyl–enzyme complex for the water and vacuum models, respectively. The largest differences observed in **Tables 6.4** and **6.5** are associated with distances O4(Ser83)···C5(MBZ) and H4(Ser83)···O6(MBZ), this is because in the starting enzyme-substrate complex **6.1** the ligand MBZ is in closer proximity to Ser83 in both water and vacuum models than in the toluene model. An explanation for this is that the two backbone hydrogen from the oxyanion hole region (Met17 and Gln84) that hold the reactant in place are further apart in the toluene model (3.80Å) than in water and vacuum models (3.38 Å and 3.29 Å). During the 25 ns MD simulation the average value of this distance is of 3.98 +/-0.33 Å in toluene and 3.48 +/-0.23 Å in water.

Table 6.4. Comparison of key distances during the formation of the acyl–enzyme from toluene model to water model. Δ is the difference between the two results (water-toluene). A threshold of 0.1 Å has established in order to differentiate between small changes (black) and bigger ones (blue).

	distances (Å)						
	O1···H2	**H2–N2**	**N3···H4**	**H4–O4**	**H4···O6**	**O4···C5**	**C5–O6**
w6.1 - t6.1	0.02	0.00	-0.06	0.00	-0.61	-0.26	0.00
wTS(6.1-6.2) - tTS(6.1-6.2)	0.05	-0.01	-0.03	-0.14	-0.14	-0.02	0.02
w6.2 - t6.2	0.05	-0.01	0.04	-0.16	-0.24	0.08	-0.02
wTS(6.2-6.3) - tTS(6.2-6.3)	0.13	-0.02	0.13	-0.16	-0.19	0.00	-0.02
w6.3 - t6.3	0.03	0.00	-0.04	-0.03	0.00	-0.01	0.00

Table 6.5. Comparison of key distances during the formation of the acyl–enzyme from toluene model to vacuum model. Δ is the difference between the two results (vacuum-toluene). A threshold of 0.1 Å has established in order to differentiate between small changes (black) and bigger ones (blue).

	distances (Å)						
	O1···H2	H2–N2	N3···H4	H4–O4	H4···O6	O4···C5	C5–O6
v6.1 - t6.1	0.07	0.00	-0.13	0.02	-0.69	-0.34	0.00
vTS(6.1-6.2) - tTS(6.1-6.2)	0.08	-0.02	0.02	-0.30	0.09	0.44	-0.03
v6.2 - t6.2	0.00	0.01	0.01	-0.07	-0.31	0.00	0.00
vTS(6.2-6.3) - tTS(6.2-6.3)	0.09	-0.01	0.17	-0.14	-0.23	-0.02	0.09
v6.3 - t6.3	0.05	0.00	-0.22	-0.07	0.02	-0.01	-0.14

The associated relative energies for the three models are presented in **Figure 6.9**. From **Figure 6.9** it can be seen that the relative energies are lower for the water and vacuum models than for the toluene model. Nevertheless, in the three cases the rate limiting step of the reaction for the acylation step is **TS(6.2-6.3)**. This step was therefore further investigated via an 'ElectroScan' and 'NegaScan' procedures in order to identify potential residues (hotspots) that could lower this barrier.

6.5 Charge modification procedures results on the acylation step

6.5.1 ElectroScan results on the rate limiting step for the acylation pathway (toluene model)

The impact of a given mutation in the enzyme on the rate limiting barrier between **t6.2** and **tTS(6.2-6.3)** was estimated by setting to zero the MM charges on the corresponding side chain and recalculation of the energy of the rate limiting barrier corresponding to stationary points. This procedure was repeated for all residues of the enzyme. The results of this 'ElectroScan' for **tTS(6.2-6.3)** of the toluene model are presented in **Figure 6.10**.

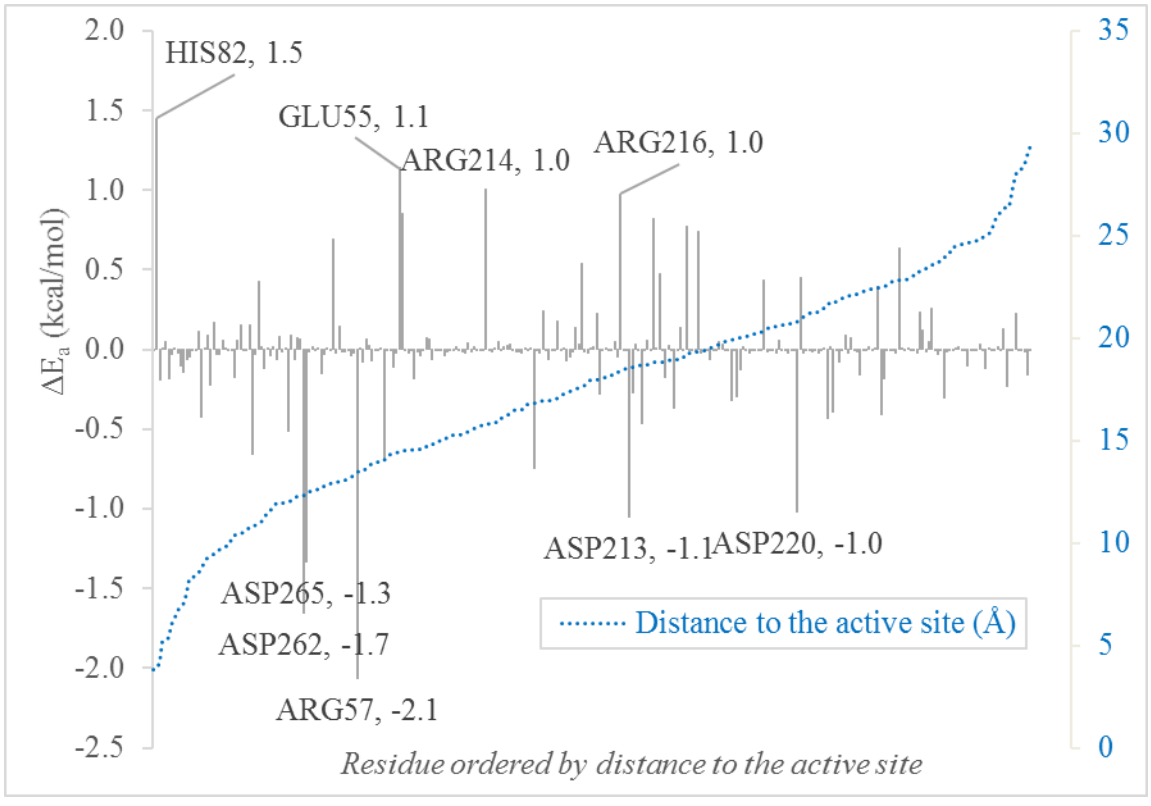


Figure 6.10. 'ElectroScan' results for **tTS(6.2-6.3)** from the toluene model.

The calculations in the toluene model identified five amino acid positions for which a charge nullification induces a significant reduction of the reaction barrier ($\Delta E < -1$ kcal/mol). These residues are Arg57 (-2.1 kcal/mol), Asp262 (-1.7 kcal/mol), Asp265 (-1.3 kcal/mol), Asp213 (-1.1 kcal/mol) and Asp220 (-1.0 kcal/mol). In these results, there is a clear preference for the identification of acidic and basic residues at hotspots, given the 20 potential gene-encoded amino acids. The same observation was also made in **Chapter 5** 'ElectroScan' results. Similarly to the transition states studied in Chapter 5, **tTS(6.2-6.3)** is associated with a proton transfer and thus particularly sensitive to a modification of the charge environment. In **Figure 6.10** residues are ordered by their distance to the active site, it is observed that the hotspot residues are not found at a specific distance from the active site.

To have a better understanding of why the charge nullification of these residues have a significant impact on the reaction barrier and why for some the barrier is increased and for other it decreases, a 3D representation of the toluene hotspot residue positions is compared to the **tTS(6.2-6.3)** reaction vector (**Figure 6.11**). Residues for which charge nullification reduces the energy barrier for this step are located in the same direction as of the proton transfer vector when they are basic and in the opposite direction when they are acidic. The reverse situation can be seen for residues that increase the barrier (orange colored residues, **Figure 6.11**). The reduction of the energy barrier due to a charge nullification results from modifying the unfavorable electrostatic potential for the proton transfer by suppressing positive charges in the same direction as the vector of the proton transfer or negative charges in the opposite direction.

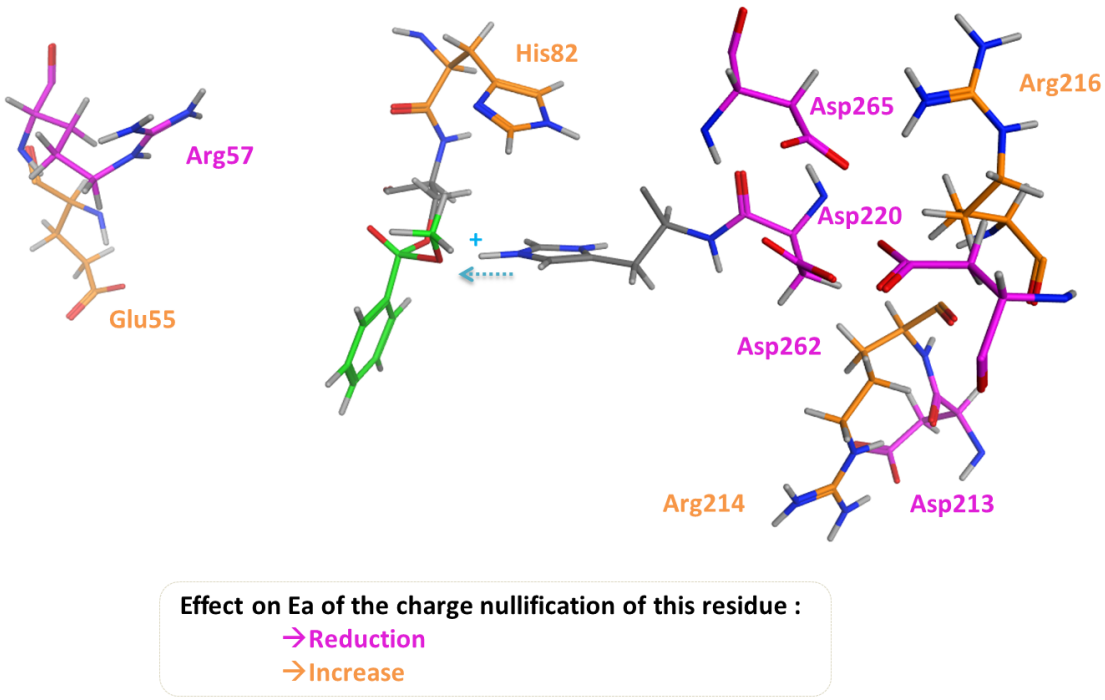


Figure 6.11. Location of the hotspot residues for the **tTS(6.2-6.3)** toluene 'ElectroScan'. The proton that is transferred during this step is indicated by a plus sign in cyan, the vector of this transfer is indicated by a cyan dotted arrow. Residues for which charge nullification reduces the energy barrier between **t6.2** and **tTS(6.2-6.3)** are represented in pink and those for which it increases this barrier are represented in orange.

6.5.2 Using ElectroScan results to suggest experimental mutations

Because it is experimentally impossible to knock out the electrostatic contribution of an amino acid the 'ElectroScan' procedure is only a preliminary study. To reproduce the effect of the 'ElectroScan' one option would be to mutate a charge residue into an uncharged residue. In order to narrow the choice, the residue conservation of the hotspot positions was examined. To do so the MSA of PBL with its 500 closest sequences within Uniprot[334] was used. To facilitate the interpretation a sequence logo representation for this

MSA was generated using WebLogo 3[336] which is presented in **Figure 6.12a**. The residue conservation between PBL and PAL (the homology model template) at these positions (**Figure 6.12b**) was also examined.

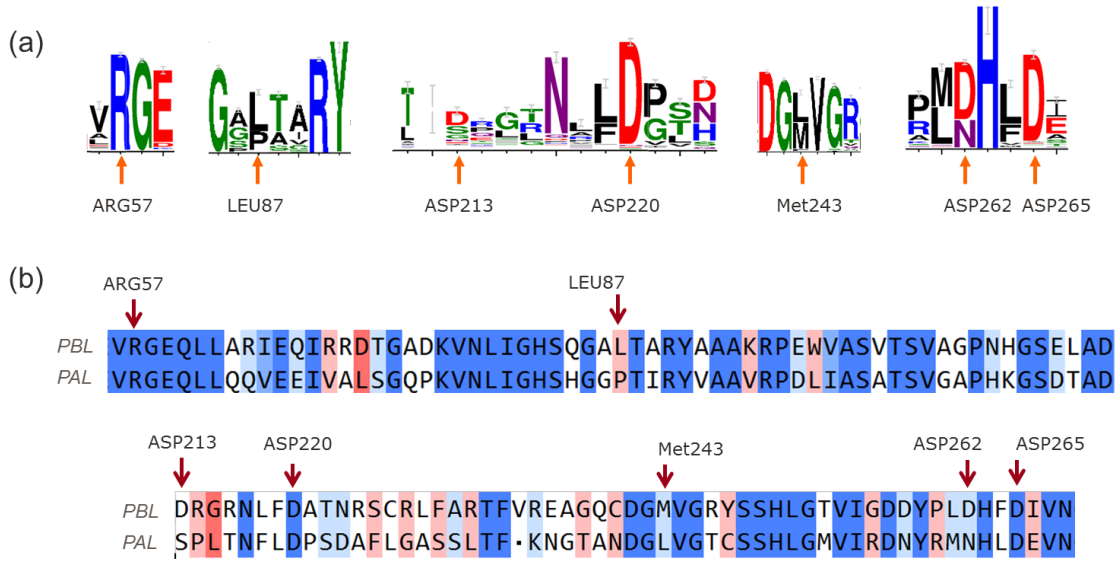


Figure 6.12. Conservation analysis of the 'ElectroScan' and 'NegaScan' hotspots (a) WebLogo representation of a MSA of the 500 closest sequences to the sequence of the PBL on Uniprot (b) Pairwise sequence alignment between PBL and PAL. Hotspots are indicated by an arrow.

Arg57 is the largest hotspot as the charge nullification of the partial charges of the side chain of this residue reduces the energy barrier between **t6.2** and **tTS(6.2-6.3)** by 2.1 kcal/mol. The MSA shows that at this position is clearly conserved as an arginine (**Figure 6.12a**). Also after inspection of the model structure it seems to be associated with a structural function with hydrogen bonds to three backbone carbonyl (Ser49, Gly16 and Pro15) and to the hydroxyl group of Thr88. Overall Arg57 does not seem an appealing residue for mutation.

The second highest hotspot is Asp262, with a predicted reduction of the reaction barrier of 1.7 kcal/mol consecutive to the charge nullification. The MSA shows that at this position an asparagine residue can also be found and in the PAL sequence this position is indeed occupied by an asparagine (**Figure 6.12**). Asparagine is structurally close to the aspartic acid and thus a D262N mutation could potentially be tolerated and should reproduce the effect of the 'ElectroScan' procedure. To explore this hypothesis, the impact of a D262N mutation was tested by *in silico* mutation; in this protocol the activation energies were re-evaluated just by performing single point calculations on a mutated intermediate-reactant complex and on a mutated transition state. The *in silico* mutation of Asp262 to an asparagine gives a ΔE_a of -1.7 kcal/mol and thus lends support to this hypothesis.

The next on the list are Asp265, Asp213 and Asp220. Asp265 and Asp220 can be ruled out because they are both strictly conserved which is not surprising as they are part of the calcium binding site. The WebLogo representation of residue Asp213 shows less conservation at this position than for other residues in the sequence, however we can observe a small conservation for aspartic acid and serine. In the PAL sequence this residue is a serine. The *in silico* mutation of Asp213 to a serine gives a ΔE_a of -1.0 kcal/mol. Nevertheless, as Asp213 forms a salt bridge to an arginine residue it could potentially have a structural role.

In order to investigate whether the 'ElectroScan' results can be combined we have calculated the *in silico* double mutant D262N/D213S which gave a ΔE_a of -2.8 kcal/mol. From this results we can see that in this case the results are complementary.

6.5.3 ElectroScan results on the rate limiting step for the acylation pathway (other models)

Table 6.6 shows the top ten residues for which the 'ElectroScan' procedure induces the larger change on the activation energy for **TS(6.2-6.3)** in toluene, water and in vacuum models. The 'ElectroScan' for **TS(6.2-6.3)** in water and in vacuum identified similar hotspots than in toluene. The intensities however seem to be larger in water and vacuum models. For example, the ΔE_a of Asp213 is -1.1 in toluene kcal/mol, -1.6 kcal/mol in water and -2.0 kcal/mol in vacuum. Calculations could thus be done with a vacuum model to gain an initial indication of the favorability of potential mutations.

Table 6.6. Residues for which the 'ElectroScan' for **TS(6.2-6.3)** induces the larger change of the activation energy (ΔE_a). Hotspots identified in the three models are represented in blue.[a]

tTS(6.2-6.3)		**wTS(6.2-6.3)**		**vTS(6.2-6.3)**	
Residue	**ΔE_a**	**Residue**	**ΔE_a**	**Residue**	**ΔE_a**
Arg57	-2.1	Arg57	-3.5	Asp262	-5.8
Asp262	-1.7	Asp262	-2.7	Arg57	-4.7
His82	1.5	Asp265	-2.2	Asp265	-2.5
Asp265	-1.3	Asp220	-2.0	Arg216	2.2
Glu55	1.1	His82	1.7	Asp213	-2.0
Asp213	-1.1	Asp213	-1.6	His82	2.0
Asp220	-1.0	Arg216	1.6	Arg235	1.9
Arg214	1.0	Arg246	1.3	Arg246	1.6
Arg216	1.0	Asp257	-1.2	Asp220	-1.5
Asp117	0.9	Asp258	-1.1	Glu114	-1.5

[a] Relative electronic activation energies (ΔE_a) in kcal/mol are given relative to the unchanged electrostatic environment.

6.5.4 NegaScan results on the rate limiting step for the acylation pathway (toluene model)

The idea of the 'NegaScan' procedure is to have a modified version of the original 'ElectroScan' procedure where not only is there a charge nullification of the partial charges of the side chains but also an electrostatic 'perturbation' is introduced by assigning a -1 charge at the Cβ that has the property of being a common side chain atom for almost all the amino acids (with the exception of Glycine). The results of this 'NegaScan' procedure applied to the step associated to **tTS(6.2-6.3)** of the toluene model is shown in **Figure 6.13**. **Table 6.7** shows the residues for which the 'NegaScan' procedure induces the larger reduction on the activation energy.

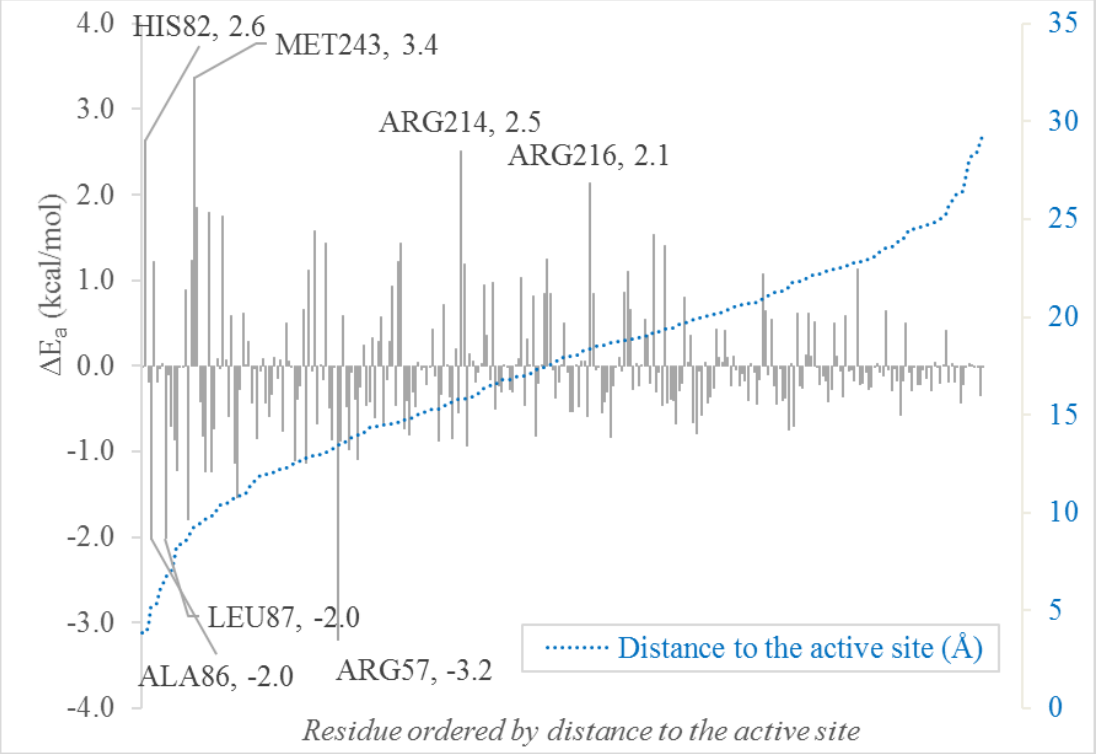


Figure 6.13. 'NegaScan' results for **tTS(6.2-6.3)** from the toluene model.

Table 6.7. Residues for which the 'NegaScan' for **tTS(6.2-6.3)** from the toluene model induces the larger change of the activation energy (ΔE_a).[a]

Residue	**ΔE_a**
Met243	3.4
Arg57	-3.2
His82	2.6
Arg214	2.5
Arg216	2.1
Ala86	-2.0
Leu87	-2.0

[a] Relative electronic activation energies (ΔE_a) in kcal/mol are given relative to the unchanged electrostatic environment.

These results demonstrate that the 'NegaScan' approach is able to explore new regions of the protein compared to the 'ElectroScan' by identifying not only acidic/basic residues but also neutral residues, such as Met243 (3.4 kcal/mol) and Leu87 (-2.0 kcal/mol).

It appears that there is a stronger distance dependence effect on the intensity of the peaks (**Figure 6.13**) than there is in the 'ElectroScan' results (**Figure 6.10**). In **Figure 6.13** residues are ordered by their distance to the active site, it is observed that the most important hotspot summarized in **Table 6.7** are all found within 20 Å of the active site.

The highest hotspot is Met243 (3.4 kcal/mol). We expect that having strongly positive ΔE_a is a sign that the introduction of a positive charge at this location (by mutating to

lysine or arginine for example) could also lead to a gain of activity. We have tested this hypothesis on Met243 and the *in silico* mutation of Met243 to a lysine gives a ΔE_a of -1.0 kcal/mol and thus lends support to this hypothesis. Nevertheless, it turns out that a saturation mutagenesis was done at this position and that M243K did not increase the activity but in fact decrease it (K. Brown, GSK, personal communication). This can be explained by the position of this residue in the binding pocket, during the molecular dynamics it can be observed that Met243 is a highly mobile residue that occasionally disturbs the binding of the ligand. It is thus not surprising that the best results from the saturation mutagenesis were mutation to smaller residues like alanine. There are three other residues in **Table 6.7** with positive ΔE_a which are His82, Arg214 and Arg216. These are already basic residues and were thus not further investigated.

To reproduce the effect of the 'NegaScan' on hotspots with strongly negative ΔE_a one needs to replace a neutral residue by a negatively charged one, aspartic acid is the easier to accommodate being smaller than glutamic acid. The potential for mutation of the hotspots in **Table 6.7** were first examined by analyzing the residue conservation at these positions (**Figure 6.12**). In a second time the hotspots were prioritized by performing *in silico* mutation to aspartic acid to evaluate if there was enough space for a mutation. The conservation analysis discarded Arg57 because it is strictly conserved. The *in silico* mutation discarded Ala86 because not enough space was available for the mutation. Finally, Leu87 is conserved to both leucine and proline, the *in silico* mutation to aspartic acid was possible and gave an activation energy reduction of 1.8 kcal/mol. L87D was thus another mutation communicated to our experimental collaborators.

6.5.5 NegaScan results on the rate limiting step for the acylation pathway (other models)

Table 6.8 shows the top fifteen residues for which the 'NegaScan' procedure induces the larger change on the activation energy for **TS(6.2-6.3)** in toluene, water and in vacuum models. The 'NegaScan' for **TS(6.2-6.3)** in water and in vacuum identified similar hotspots than in toluene. The intensities however seem to be larger in water and vacuum models. For examples ΔE_a for Leu87 is -2.0 in toluene, -2.4 in water and -3.1 in vacuum. As for the 'ElectroScan', calculations could thus be done with an vacuum model but with bearing in mind that the intensities of the results are changed.

Table 6.8. Residues for which the 'NegaScan' for **TS(6.2-6.3)** induces the larger change of the activation energy (ΔE_a). Hotspots identified in the three models are represented in blue.[a]

tTS(6.2-6.3)		**wTS(6.2-6.3)**		**vTS(6.2-6.3)**	
Residue	**ΔE_a**	**Residue**	**ΔE_a**	**Residue**	**ΔE_a**
Met243	3.4	Arg57	-5.6	Arg57	-7.5
Arg57	-3.2	Ala107	4.3	Met243	6.2
His82	2.6	Val244	4.3	Ala107	5.7
Arg214	2.5	Met243	3.7	Val244	4.4
Arg216	2.1	Arg216	3.4	Arg246	4.3
Ala86	-2.0	Arg246	3.3	Arg216	4.3
Leu87	-2.0	Cys240	3.2	Phe264	3.9
Phe264	1.9	Ile266	3.0	Thr206	3.7
Thr88	-1.8	Thr206	2.9	Cys240	3.6
Ser204	1.8	Arg214	2.5	Ile266	3.5
Ile266	1.8	Thr88	-2.5	Arg214	3.4
Arg246	1.6	Ser204	2.5	Thr88	-3.3
Arg90	-1.5	His82	2.4	Ser204	3.2
Arg227	1.5	Leu261	2.4	Arg235	3.1
Leu261	1.4	Leu87	-2.4	Leu87	-3.1

[a] Relative electronic activation energies (ΔE_a) in kcal/mol are given relative to the unchanged electrostatic environment.

6.5.6 Effect of the selected mutants on the stability of the product of the acylation step

Given the endothermicity of the acylation step (**Figure 6.9**) we have also investigated the effect of D262N, D213S, and L87D mutations on the barrier between **tTS(6.2-6.3)** and **t6.3**. This barrier is initially 28.3 kcal/mol and the effect of the *in silico* mutations on this barrier is reported in **Table 6.9**. These results show that all these mutations increase the **tTS(6.2-6.3)**/**t6.3** barrier and as we already know that the same mutations decrease the **t6.2**/**tTS(6.2-6.3)** barrier it means that these mutations could also make the reaction less endothermic.

Table 6.9. Effect of various *in silico* mutations on the energy barrier between **tTS(6.2-6.3)** and **t6.3** from the toluene model.[a]

***In silico* mutation**	**ΔΔE**
D262N	4.1
D213S	3.0
L87D	2.7

[a] Relative electronic energies (ΔΔE) in kcal/mol are given relative to the original energy barrier between **tTS(6.2-6.3)** and **t6.3** in the wild type system.

6.6 Reaction mechanism description for the amidation step

Given the capacity of the calculation with the vacuum model to identify the same rate limiting step and 'ElectroScan' hotspots than the calculations with the toluene model for the acylation step, the amidation step was studied using the vacuum model.

The investigation of the amidation reaction from the vacuum model reveals that amidation step requires a two-step mechanism: the nucleophilic attack of a second ligand amine nitrogen to the acyl-enzyme complex and the formation of an amide product with regeneration of the active site (**Figure 6.14a**). A schematic representation of the two transition state of the reaction (**vTS(6.4-6.3)** and **vTS(6.5-6.6)**) including only the closest atoms around substrate, is given in **Figure 6.14b-c**. **Table 6.10** shows the relatives energies and the most important distances during the amidation reaction.

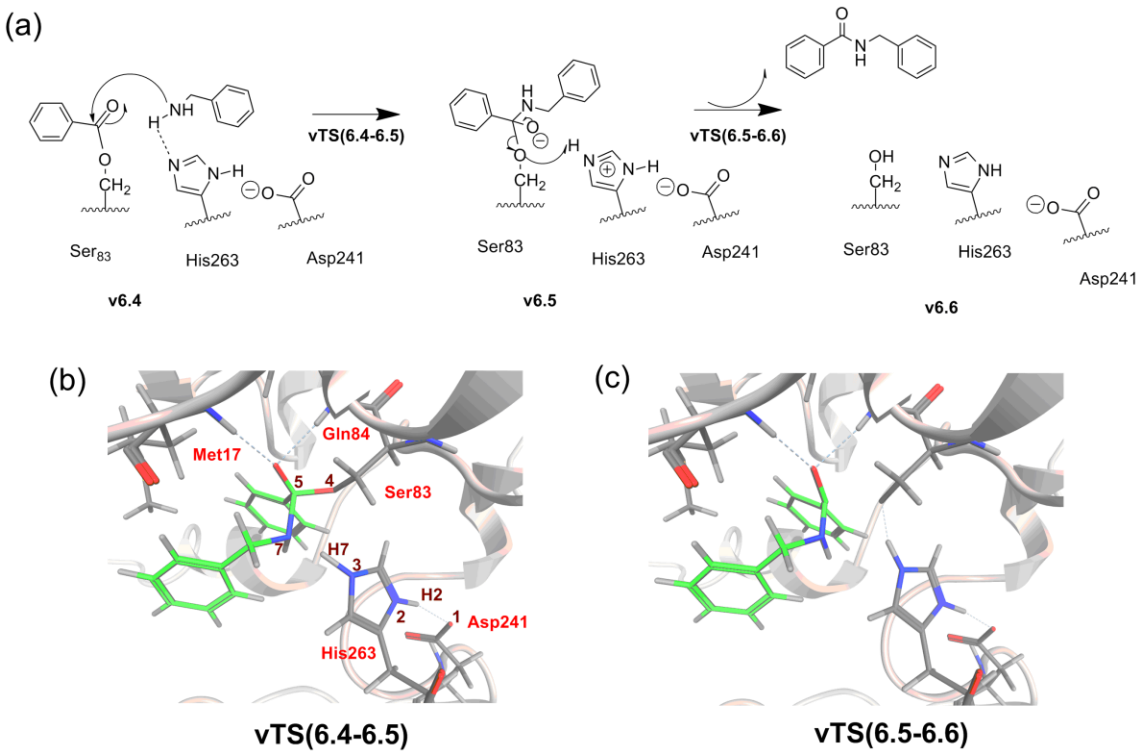


Figure 6.14. Schematic representation of the amidation pathway and the corresponding transition states structures in vacuum.

Table 6.10. Relative energies and key distances during the amidation reaction.[a, b]

	ΔE (kcal/mol)	distances (Å)						
		O1⋯H2	**H2–N2**	**N3⋯H7**	**H7⋯O4**	**H7–N7**	**O4–C5**	**C5⋯N7**
v6.4	0	1.73	1.06	2.01	2.91	1.02	1.35	2.63
vTS(6.4-6.5)	15.2	1.72	1.07	1.27	2.47	1.35	1.47	1.66
v6.5	11.9	1.60	1.09	1.06	2.49	1.73	1.51	1.57
vTS(6.5-6.6)	13.4	1.49	1.11	1.10	1.53	2.38	2.04	1.43
v6.6	-3.4	1.64	1.07	1.77	1.00	2.62	2.76	1.37

[a] Electronic energies (ΔE) in kcal/mol are given relative to the energy of **v6.4**.

[b] Results obtained using B3LYP/6-31G*/OPLS2005 as level of theory

Once the formation of the acyl–enzyme complex is completed and an amine molecule is available in the active site region, the amidation reaction can take place. In this reaction the alcohol group of the ester formed within the acyl–enzyme complex is exchanged by an amine group.

The first step of the amidation involves the attack of the nitrogen of the second ligand, phenylmethanamine (PMA), on the acyl complex. In the optimized enzyme-substrate complex **v6.4** PMA is stabilized between the acyl complex and His263. One of the hydrogens of the amine group of PMA is pointing toward the Nε atom of His263 establishing a hydrogen bond (N3(His263)•••H7(PMA) distance of 2.01 Å). Such interaction allows the approach of the nitrogen atom near carbon C5 of the acyl–enzyme complex (N7(PMA)•••C5(MBZ) distance of 2.63 Å) promoting in this way the amidation process. In the first transition state **vTS(6.4-6.5**) (characterized by an imaginary frequency of -1158 cm^{-1}) the amine group of PMA is approaching carbon C5 of the acyl–enzyme complex (N7(PMA)•••C5(MBZ) distance of 1.66 Å), and the hydrogen atom is shared with one of the nitrogen atoms of His263 (N7(PMA)–H7(PMA): 1.35 Å; and H7(PMA)–N3(His263): 1.27 Å). The hydrogen bond between Asp241 and the Nδ atom of His263 is unchanged. Following **vTS(6.4-6.5**) the second tetrahedral intermediate is formed. In the first intermediate **v6.5**, the amine molecule becomes covalently bonded to the acyl–enzyme complex forming a tetrahedral intermediate that is negatively charged. The hydrogen from the amine group of PMA is completely transferred toward His263 (H7(PMA)–N3(His263) bond length of 1.06), this nitrogen atom is thus positively

charged. This first step requires an activation energy of 15.2 kcal/mol, and a change of the overall energy of reaction of 11.9 kcal/mol (**Table 6.10**).

The second and final step of the amidation mechanism involves the release of the amide product and turnover of the active site. In the second transition state **vTS(6.5-6.6**) (characterized by an imaginary frequency of -156 cm^{-1}) Ser83 is partially dissociated from the substrate (2.04 versus 1.51 Å in **v6.5**). The proton that was previously attached to His263 is now shared with Ser83 (distance H7(His263)–O4(Ser83) of 1.53 Å and distance H7(His263)–N3(His263) of 1.10 Å). There is also a shorter distance between Asp241 and His263 (O1(Asp241)–H2(His263) is now 1.49 Å was 1.60 Å in the **v6.5** structure). In the product of the reaction **v6.6**, Ser83 is finally reprotonated and unbound from the product. The product is now free to dissociate from the active site, although the carbonyl group remains stabilized by hydrogen bond interactions with the oxyanion hole. This final reaction requires an activation energy of 1.5 kcal/mol relative to the previous intermediate and is exergonic by 15.3 kcal/mol, relative to **v6.5** (**Table 6.10**). The overall barrier for the amidation reaction is 15.2 kcal/mol. Furthermore, the product of this reaction is formed in an overall exothermic reaction of -3.4 kcal/mol.

The complete energy path from the vacuum model is presented in **Figure 6.15**. The rate limiting steps of the reaction are **vTS(6.2-6.3)** for the acylation step and **vTS(6.4-6.5)** for the amidation step.

The rate-determining step in the acylation pathway was previously investigated by 'ElectroScan' and 'NegaScan' procedures from which D262N/D213S and L87D

mutations were suggested, respectively. To study the effect of these mutations on the rate-determining step of the amidation pathway this was further investigated via the same procedures.

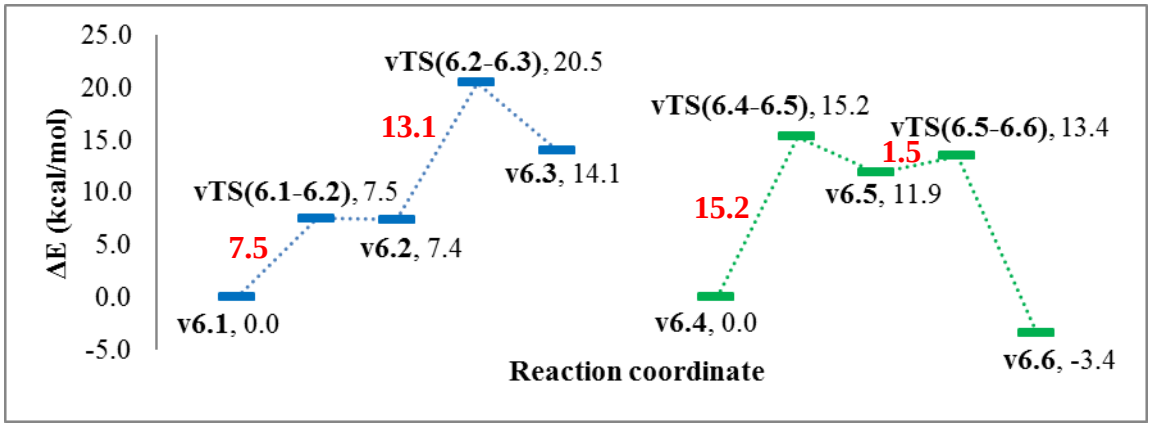


Figure 6.15. Potential energy profile involved in the formation of the acyl–enzyme complex (blue) and the amidation (green) obtained from the vacuum model. QM/MM electronic energies (ΔE) in kcal/mol are given relative to the reactant state for each step.

6.7 Charge modification procedures results on the amidation step

6.7.1 ElectroScan results on the rate limiting step for the amidation pathway (vacuum model)

The impact of a given mutation in the enzyme on the rate limiting barrier between **v6.4** and **vTS(6.4-6.5**) was estimated by applying the 'ElectroScan' procedure which was repeated for all residues of the enzyme. The results are presented in **Figure 6.16**. From these results, it can be seen that hotspot residues are shared between **vTS(6.2-6.3)** and **vTS(6.4-6.5**) but with the opposite effect; these are Asp265 (-5.8 for **vTS(6.2-6.3)** and 6.5 for **vTS(6.4-6.5**)) and Asp213 (-2.0 for **vTS(6.2-6.3)** and 1.8 for **vTS(6.4-6.5**)) for example.

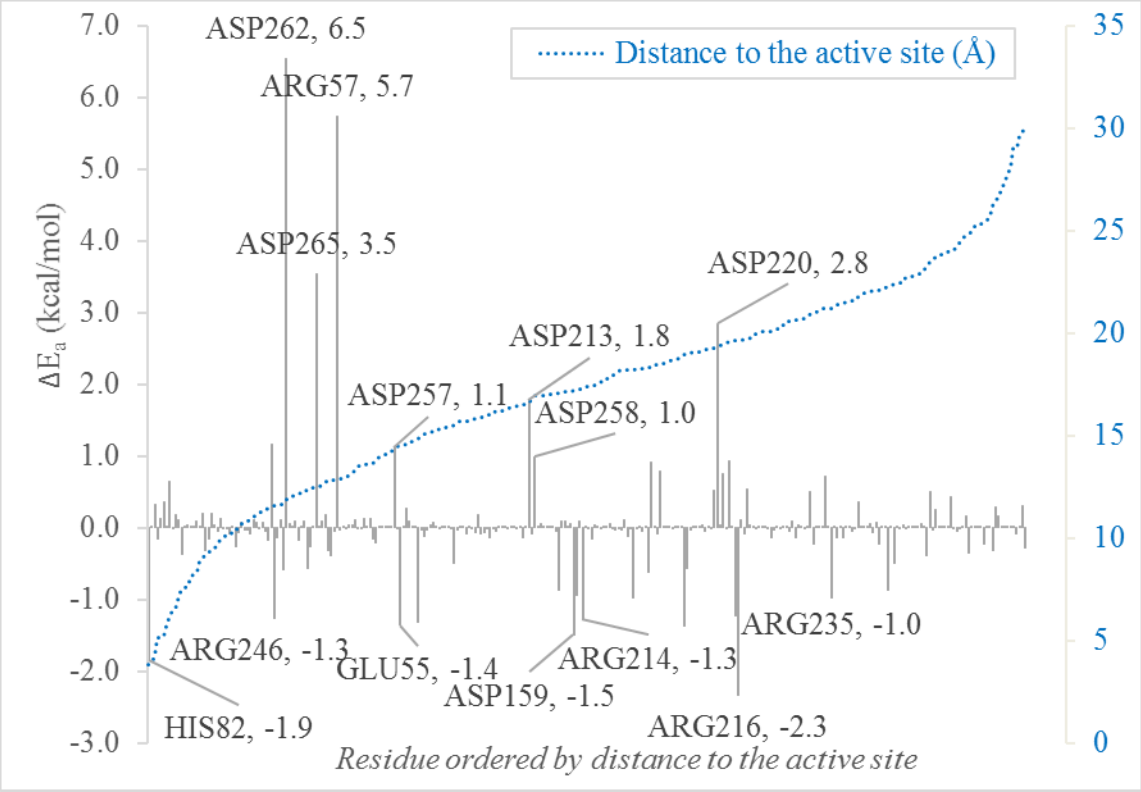


Figure 6.16. 'ElectroScan' results for **vTS(6.4-6.5)** from the vacuum model.

The opposite effect of these results can be linked to the reaction mechanism. As summarized in **Figure 6.17**, all the steps are associated with a proton transfer. More precisely some proton transfers like **vTS(6.2-6.3)/vTS(6.4-6.5)** and **vTS(6.1-6.2)/vTS(6.5-6.6)** share the same line but are in opposite directions as indicated by the arrows in **Figure 6.17**.

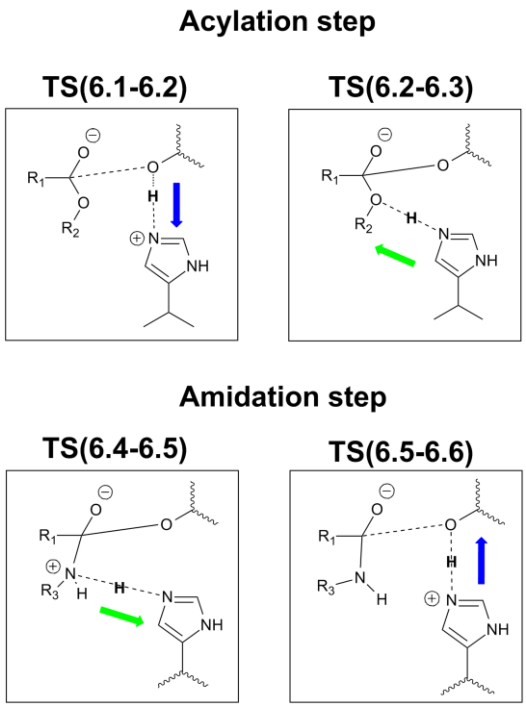


Figure 6.17. Schematic representation of proton transfers during the enzymatic amide bond formation. Proton transfer sharing the same line are represented by the same color and the direction is indicated by an arrow.

The opposite nature of the effect of the 'ElectroScan' procedure of the different steps highlights the need for a multistep parameterization strategy for our system. Therefore, rather than identifying hotspots for one specific step another option would be to rank residues according to the sum of their effect on different steps. This approach was taken for steps associated with **vTS(6.2-6.3)** and **vTS(6.4-6.5**) and results for the five residues with the largest magnitudes sum are summarized in **Table 6.11**. From these results we can see once again that for almost all the residues in this table 'ElectroScan' results are of opposite sign. The only exception is residue Glu114 for which both results are negatives (-1.5 for **vTS(6.2-6.3)** and for -0.6 **vTS(6.4-6.5**)). However, the intensities of these

'ElectroScan' results are low for hotspots obtained using the vacuum model – the 'ElectroScan' result for Glu114 of **tTS(6.2-6.3)** in toluene was only -0.5 kcal/mol. Additionally, Glu114 forms a salt bridge to an arginine residue and could potentially have a structural role, which is further supported by the strict conservation of an acidic residue (E/D) at this position, from the multiple sequence alignment. Finally, Glu114 has been submitted to saturation mutagenesis and none of the mutations increase the activity of the enzyme on the studied reaction, instead most of the mutations almost completely knock-out the enzymatic activity (K. Brown, GSK, personal communication).

Table 6.11. 'ElectroScan' results for **vTS(6.2-6.3)** and **vTS(6.4-6.5**) from the vacuum model for residues with the larger magnitude absolute sums.[a]

	ΔE_a		
Residue	**vTS(6.2-6.3)**	**vTS(6.4-6.5)**	**Absolute sum**
Glu114	-1.5	-0.6	2.0
Asp220	-1.5	2.8	1.3
Arg57	-4.7	5.7	1.0
Asp265	-2.5	3.5	1.0
Arg235	1.9	-1.0	1.0
Asp117	0.5	-1.3	0.8
Arg224	0.6	-1.4	0.8
Asp262	-5.8	6.5	0.8
Glu170	0.4	-0.9	0.5

[a] Relative electronic activation energies (ΔE_a) in kcal/mol are given relative to the unchanged electrostatic environment.

6.7.2 NegaScan results on the rate limiting step for the amidation pathway (vacuum model)

The results of the 'NegaScan' procedure applied to the step associated to **vTS(6.2-6.3)** of the vacuum model is shown in **Figure 6.18**. It appears from these results that as for the 'ElectroScan' hotspot residues are shared between **vTS(6.2-6.3)** and **vTS(6.4-6.5**) but with the opposite effect; these are Met243 (6.2 for **vTS(6.2-6.3)** and -6.7 for **vTS(6.4-6.5**)) and Leu87 (-3.1 for **vTS(6.2-6.3)** and 3.3 for **vTS(6.4-6.5**)) for example. We have

thus also explored the sum of the two 'NegaScan' procedure and results are presented in **Table 6.12**. In this table all the results are of opposite sign. For two residues however the effect on **vTS(6.2-6.3)** is really small compared to the effect on **vTS(6.4-6.5**), these are Ser113 (-0.1 for **vTS(6.2-6.3)** and for 1.6 **vTS(6.4-6.5**)) and Tyr118 (-0.1 for **vTS(6.2-6.3)** and for 1.4 **vTS(6.4-6.5**)). In the 'NegaScan' results on **vTS(6.2-6.3)** of the vacuum model Asn223 had a result of 1.5 which yield to a result of 1.2 for the toluene model, the introduction of a positive charge at this position could thus have a positive impact on the barrier. Nevertheless, Ser113 is part of the binding site and a mutation to a bulkier residues lysine or arginine could disturb ligand binding. As for Tyr118 there is already an arginine residue next to it so introduction of a positively charge residue could induce repulsion between the two residues.

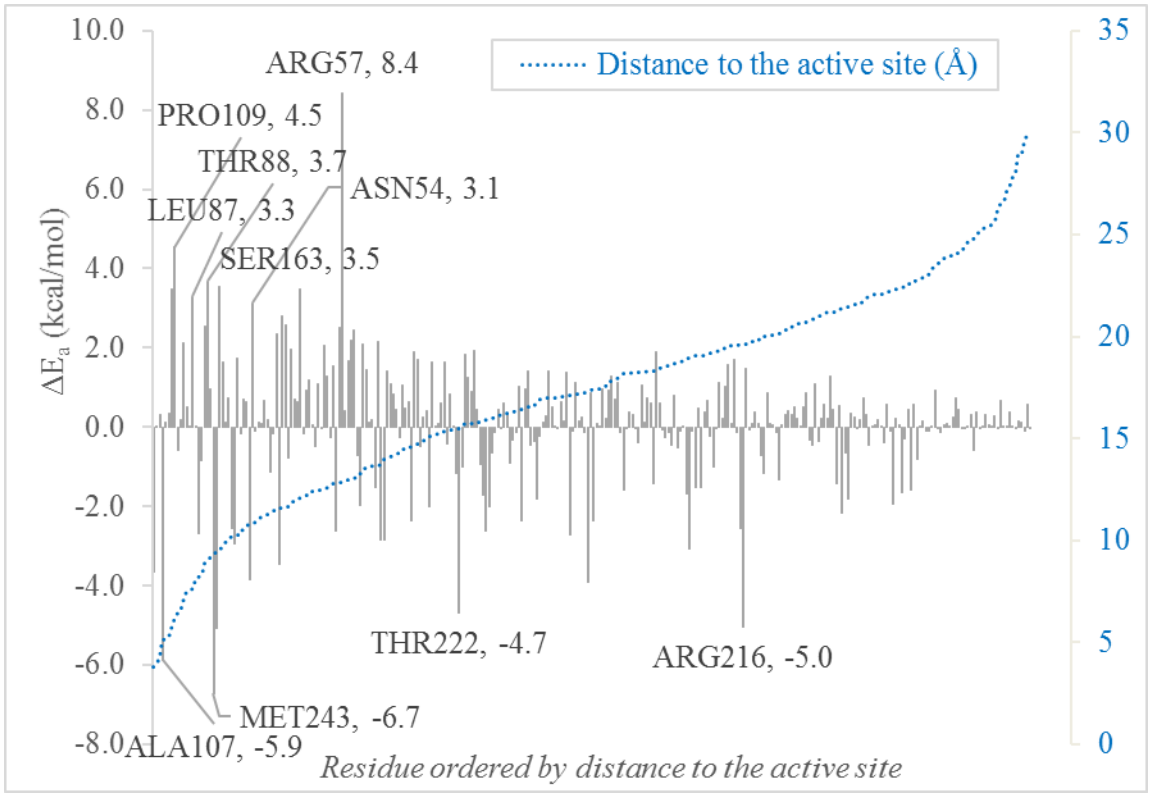


Figure 6.18. 'NegaScan' results for **vTS(6.4-6.5**) from the vacuum model.

Table 6.12. 'NegaScan' results for **vTS(6.2-6.3)** and **vTS(6.4-6.5**) from the vacuum model for residues with the larger magnitude sums.[a]

	ΔE_a		
Residue	**vTS(6.2-6.3)**	**vTS(6.4-6.5**)	**Sum**
Thr222	2.2	-4.7	2.5
Pro109	-2.1	4.5	2.4
Arg235	3.1	-1.4	1.7
Val244	4.4	-2.7	1.7
Arg224	1.5	-3.1	1.6
Ser113	-0.1	1.6	1.6
Leu18	-2.1	0.7	1.4
Tyr118	-0.1	1.4	1.3
Ser225	1.4	-2.6	1.2
Phe264	3.9	-5.1	1.2

[a] Relative electronic activation energies (ΔE_a) in kcal/mol are given relative to the unchanged electrostatic environment.

6.8 Conclusion

The goal of the study was to further investigate the QM/MM-based hotspot identification protocol established in Chapter 5. For this, an amide bond formation reaction catalyzed by a lipase from *Pseudomonas batumici* (PBL) was used as case study.

An homology model of PBL was first built using the crystallographic structure of a lipase from *Pseudomonas aeruginosa* as template. The underlying mechanism of the reaction was then investigated using QM/MM. A catalytic mechanism was suggested with a two-step acylation followed by a two-step amidation.

The 'ElectroScan' technique was then applied to the rate-limiting step of the acylation step. Partial charges of single residues side chains were set to zero and the activation energy re-evaluated in the new environments. The results gave five hotspots (with $\Delta E_a <$ -1 kcal/mol) in Arg57, Asp262, Asp265, Asp213 and Asp220. After finding these five hotspots it was studied what the best mutation to perform might be. Multiple sequence alignments were created and examined to determine which mutations to study. The impact of the selected mutations was tested by *in silico* mutation. In this protocol the activation energies were re-evaluated just by performing single point calculations on a mutated enzyme-intermediate complex and on a mutated transition state. This double single point protocol allows a quick evaluation as it considerably faster than a full mutation simulation. This procedure identified D262N and D213S as beneficial mutations.

The ‘NegaScan’ protocol was tested for the first time on the same reaction and L87D was identified as a beneficial mutation. ‘NegaScan’ could be a possible strategy to introduce new favorable electrostatic interaction in a protein.

The effect of the solvent (toluene, water and vacuum) on the acylation step mechanistic and ‘ElectroScan’ result where also investigated. It was found that with the three solvents the same mechanism was found and the similar hotspots were identified.

The subsequent application of the ‘ElectroScan’ technique to the rate-limiting step of the amidation step (the first transition state of the amidation step) revealed that the applicability of this methodology for this system was found to be limited by the reaction mechanism. More precisely, **vTS(6.2-6.3)** and **vTS(6.4-6.5**) are associated with proton transfer reactions that share the same line but in opposite direction, thus depending on the direction of the proton transfer different amino acids are identified and side chains charge nullification effects are mutually canceled.

7. General conclusion and recommendations

7.1. Conclusion

The drug discovery and development process is long and costly. New approaches that offer the potential to save time and money are thus actively pursued within the pharmaceutical industry. The work presented in this thesis has focused on addressing these issues in the context of biocatalysis development – a specific step on the drug discovery and development pipeline (**Figure 7.1**).

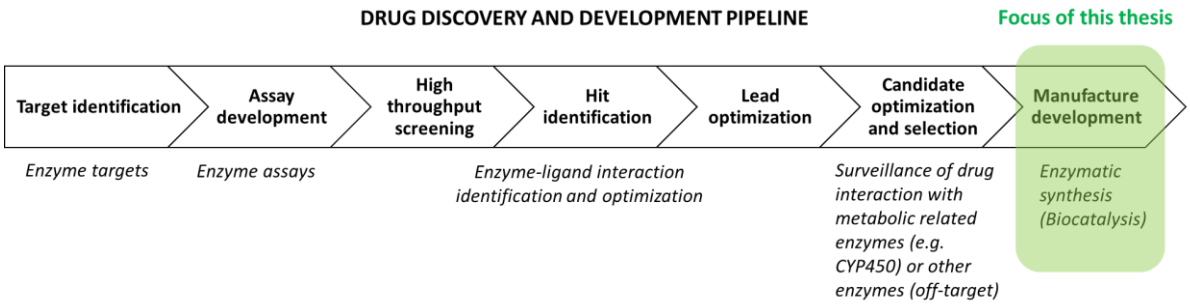


Figure 7.1. Example of enzyme-ligand systems along the drug discovery and development pipeline. The focus of the work of this thesis is shaded in green.

The motivation behind this work originated from the possibility offered by QM/MM methods to evaluate reaction energy barriers which could be used for the identification of hotspots. One of the challenges faced at the beginning of this work was the absence of "in-house" validated QM/MM methodology. The work on aldose reductase presented in **Chapter 4** has thus helped this validation. Factors such as the size of the QM region and the choice of QM methods were explored and were found critical to obtain accurate

results. This step has also created the solid foundation for other applications of QM/MM methodology along the drug discovery and development pipeline as described in **Figure 7.1**.

Standard QM/MM methods are computationally expensive and thus for a practical application approximations need to be introduced. Therefore, the methodology validated in **Chapter 4** was applied in the next step of the study, which was to investigate which approximations could be introduced to have a fast, but sufficiently accurate, screening method. In **Chapter 5**, a QM/MM protocol for hotspot identification was thus elaborated and tested. The protocol comprises a 1.2 ns NPT molecular dynamic simulation of the Michaelis complex for snapshot generation, a QM/MM mechanistic study (B3LYP/6-31G*/OPLS2005) to identify the rate-limiting step and associated stationary point structures and a finally a charge modification procedure (with either a charge deletion or charge introduction). To speed up the calculations three main approximations were introduced. First, the strategy of investigating only one snapshot from the molecular dynamics simulation was taken. We believe this is an acceptable approximation as the mechanism should remain the same. To explore the limits of this approximation the results from additional snapshots could be investigated.

Second, to speed up the energy barrier evaluation step only potential energies, rather than free energies, were calculated. Entropic contributions are thus not included in these calculations. While entropic and thermal contributions can play a significant role in determining transition state energies, in this case we are looking at relative effects and we do not expect the inclusion of entropic effects to be essential. The importance of entropic

effects could be evaluated by calculating the free energy profile of the reaction and would certainly be recommended in a more extensive mechanistic study of the enzyme.

Third, we chose to neglect possible reorganization effects due to the introduction of a mutation in the enzyme structure. This is potentially the most significant approximation in that an individual mutation could cause the collapse of the secondary structure of the enzyme and significantly alter the active site structure. Nonetheless, re-equilibrating and preparing the structure of each new system is a time-critical point in the protocol and therefore, it was not considered a viable option for a fast screening method. To minimise the probability of mutations causing a significant change in the structure of the enzyme the additional step of evaluation of the QM/MM identified hotspots via inspection of multi sequence alignment from the closely related sequences was introduced. Additionally, *in silico* mutations are carried out (on the static structures) to evaluate if enough space is available to accommodate a given mutation.

The ambiguity associated with using a non-equilibrated structure to estimate the reaction profile of mutated structures may be able to be removed in the near future. As discussed in the **Chapter 1**, under the influence of Moore's Law molecular dynamics simulations are destined to become faster and find even broader application. Therefore, a future possibility that could be explored would be the addition of a (relatively short) molecular dynamics simulation following the mutation step in our current protocol, to directly evaluate the effect of the mutation on the structure.

Computational studies of enzymes require an initial 3D structure, however protein structures are rare.[347] A large gap exists between the number of protein sequence available (greater than 100 million) and the number of experimentally solved protein structures (approximately 100,000).[347] The work presented in **Chapter 6** addresses this limitation by using a homology model strategy. This strategy seems reasonable in our study as a sufficiently close experimentally determined structure was available as template (44.6% identity). However, in cases where the sequence identity between the query and the sequence is not high enough (*e.g.*, < 30%), less confidence can be given to the predicted structure and less efficient protein design methods have to be used (*i.e.*, epPCR). However, the field of structure prediction is progressing particularly in *ab initio* type prediction – prediction without template information. The Critical Assessment of methods of protein structure prediction (CASP) is a blind test competition that evaluates current capabilities of Structure Prediction methods. Since CASP 11 (held in 2014) a dramatic improvement in *ab initio* predictions has been observed.[348-349] Krzysztof Fidelis, member of the CASP organising committee since its establishment in 1994, has recently stated that for most proteins we will soon be able to declare protein folding problem solved.[348] This could greatly facilitate enzyme (re)design.

One of the most significant time costs in biomolecular simulations is related to the presence of solvent. In **Chapter 6** an approach to reduce this cost was explored by investigating the results obtained for reaction mechanism studies and hotspot identification procedures run without solvent. These results were compared to those with water and toluene solvents. The results show that mechanistic investigation and hotspot

identification obtained without solvent gave similar results from those obtained with water or toluene solvent. Nonetheless, as high energy barriers were obtained in the calculation with the toluene solvent these results should be taken with care. This is not necessarily a problem as the same mechanism was obtained, however to strengthen these findings, additional work will need to be carried out on the toluene mechanism, in particular prudent future steps include the use of multiple snapshots to better average the potential energy surface obtained, or, alternatively, carrying out QM/MM MD simulations to obtain a free energy profile.

Overall, a QM/MM protocol for hotspot identification has been established and tested starting from both a crystallographic structure and a homology model. The results of the established QM/MM protocol run without the solvent strategy has been evaluated and seems a reasonable approximation but some further investigations on the mechanism in toluene solvent need to be done to validate the approximation. This work has thus paved the way to the establishment of an in-house QM/MM protocol for biocatalysis hotspot identification.

7.2. Recommendations

As described in Section 4.2, aldose reductase is a potential target for the treatment of tissue-based pathologies associated with diabetes mellitus complications. The results from the mechanistic study of AR in **Chapter 4** could thus be used as starting point for the development of an inhibitor via the investigation of transition-state analogues.[350-351]

A mechanistic study for amide bond formation catalyzed by a carboxylesterase was established in **Chapter 5**. To validate the proposed mechanism free energy calculations of the energy profile could be made for example by QM/MM MD simulations with umbrella sampling.[352] The stationary points identified in **Chapter 5** could be a good starting point for such studies. However, free energy calculations are not currently handled by Qsite so another software should be used and benchmarked. In this regard, Terachem[353] seems to be a particularly appealing platform as it comprisse a GPU parallelized electronic structure calculation code which could allow high speed calculations.

The QM/MM protocol described in **Chapter 5** is currently run as separate steps (apart from the charge modification procedures that have been fully automated via a python script). The next step would thus be to automate the protocol through the use of workflow tools (*e.g.*, KNIME).[354-355] A particularly interesting recent development is the automated transition state search workflow designed by the Schrodinger team which if combined with Qsite has the potential to automate the rate-limiting step identification work.[356] The utility of this approach should thus be explored by comparison against the transition states identified in the various systems studied in this work.

investigates the amide bond formation catalysed by a lipase

In **Chapter 6** the amide bond formation reaction catalysed by a lipase was investigated in different solvents. The simulations in toluene solvent did not include any water molecule in the system. However, it has been reported that a shell of water on the surface can be essential for the enzyme function notably to retain the three dimensional structure and promote sufficient conformational flexibility which are both essential for the activity of the enzyme.[357] In future work new simulation could be done by inclusion of such layer of water by for example keeping crystal water.[358] In the study described in Chapter 6 this was not possible as calculations started from an homology model.

References

1. Munos, B., Lessons from 60 years of pharmaceutical innovation. *Nature reviews. Drug discovery* **2009,** *8* (12), 959-68.

2. Plenge, R. M., Disciplined approach to drug discovery and early development. *Science Translational Medicine* **2016,** *8* (349), 349ps15-349ps15.

3. Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L., How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature reviews. Drug discovery* **2010,** *9* (3), 203-14.

4. Eder, J.; Herrling, P. L., Trends in Modern Drug Discovery. *Handbook of experimental pharmacology* **2016,** *232*, 3-22.

5. Constable, D. J. C.; Dunn, P. J.; Hayler, J. D.; Humphrey, G. R.; Leazer, J. J. L.; Linderman, R. J.; Lorenz, K.; Manley, J.; Pearlman, B. A.; Wells, A.; Zaks, A.; Zhang, T. Y., Key green chemistry research areas-a perspective from pharmaceutical manufacturers. *Green Chemistry* **2007,** *9* (5), 411-420.

6. Sheldon, R. A.; Woodley, J. M., Role of Biocatalysis in Sustainable Chemistry. *Chemical reviews* **2018,** *118* (2), 801-838.

7. Truppo, M. D., Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Medicinal Chemistry Letters* **2017,** *8* (5), 476-480.

8. National Institute of Biomedical Imaging and Bioengineering. https://www.nibib.nih.gov/science-education/science-topics/computational-modeling.

9. van de Waterbeemd, H.; Carter, R. E.; Grassy, G.; Kubinyi, H.; Martin, Y. C.; Tute, M. S.; Willett, P., Glossary of terms used in computational drug design (IUPAC Recommendations 1997). In *Pure and Applied Chemistry*, 1997; Vol. 69, p 1137.

10. van Vlijmen, H.; Desjarlais, R. L.; Mirzadegan, T., Computational chemistry at Janssen. *Journal of computer-aided molecular design* **2017,** *31* (3), 267-273.

11. Muegge, I.; Bergner, A.; Kriegl, J. M., Computer-aided drug design at Boehringer Ingelheim. *Journal of computer-aided molecular design* **2017,** *31* (3), 275-285.

12. Kuhn, B.; Guba, W.; Hert, J.; Banner, D.; Bissantz, C.; Ceccarelli, S.; Haap, W.; Korner, M.; Kuglstatter, A.; Lerner, C.; Mattei, P.; Neidhart, W.; Pinard, E.; Rudolph, M. G.; Schulz-Gasch, T.; Woltering, T.; Stahl, M., A Real-World Perspective on Molecular Design. *Journal of medicinal chemistry* **2016,** *59* (9), 4087-102.

13. Hillisch, A.; Heinrich, N.; Wild, H., Computational Chemistry in the Pharmaceutical Industry: From Childhood to Adolescence. *ChemMedChem* **2015,** *10* (12), 1958-62.

14. Feng, J. A.; Aliagas, I.; Bergeron, P.; Blaney, J. M.; Bradley, E. K.; Koehler, M. F.; Lee, M. L.; Ortwine, D. F.; Tsui, V.; Wu, J.; Gobbi, A., An integrated suite of modeling tools that empower scientists in structure- and property-based drug design. *Journal of computer-aided molecular design* **2015,** *29* (6), 511-23.

15. Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R., Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation

Protocol and Force Field. *Journal of the American Chemical Society* **2015,** *137* (7), 2695-2703.

16. Moore, G. E., Cramming More Components Onto Integrated Circuits, Electronics,(38) 8. April: 1965.

17. Trist, I. M. L.; Botta, M.; Fallacara, A. L., Chapter 11 Application of Molecular Modelling to Speed-up the Lead Discovery Process. In *Computational Tools for Chemical Biology*, The Royal Society of Chemistry: 2018; pp 281-316.

18. Götz, A. W.; Wölfle, T.; Walker, R. C., Quantum Chemistry on Graphics Processing Units. 2010; pp 21-35.

19. Allen, D., Where will we get the next generation of medicinal chemists? *Drug discovery today* **2016,** *21* (5), 704-6.

20. Dréanic, M.-P.; Edge, C. M.; Tuttle, T., New Insights into the Catalytic Mechanism of Aldose Reductase: A QM/MM Study. *ACS Omega* **2017,** *2* (9), 5737-5747.

21. Muller, P., Glossary of terms used in physical organic chemistry (IUPAC Recommendations 1994). *Pure Appl. Chem.* **1994,** *66*, 1077–1184.

22. Berg JM, T. J., Stryer L., *Biochemistry*. 5th edition. New York: W H Freeman: 2002.

23. McDonald, A. G.; Boyce, S.; Moss, G. P.; Dixon, H. B. F.; Tipton, K. F., ExplorEnz: a MySQL database of the IUBMB enzyme nomenclature. *BMC Biochemistry* **2007,** *8*, 14-14.

24. IUBMB Enzyme Nomenclature. http://www.sbcs.qmul.ac.uk/iubmb/enzyme/.

25. Gutteridge, A.; Thornton, J. M., Understanding nature's catalytic toolkit. *Trends Biochem Sci* **2005,** *30* (11), 622-9.

26. Pauling, L.; Corey, R. B.; Branson, H. R., The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences* **1951,** *37* (4), 205.

27. Pauling, L.; Corey, R. B., Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds: Two New Pleated Sheets. *Proceedings of the National Academy of Sciences of the United States of America* **1951,** *37* (11), 729-740.

28. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M., The protein data bank: A computer-based archival file for macromolecular structures. *Journal of molecular biology* **1977,** *112* (3), 535-542.

29. Micsonai, A.; Wien, F.; Kernya, L.; Lee, Y.-H.; Goto, Y.; Réfrégiers, M.; Kardos, J., Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences* **2015,** *112* (24), E3095.

30. Rose, G. D.; Glerasch, L. M.; Smith, J. A., Turns in peptides and proteins. In *Advances in protein chemistry*, Elsevier: 1985; Vol. 37, pp 1-109.

31. Venkatachalam, C. M., Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers* **1968,** *6* (10), 1425-1436.

32. Schlick, T., *Molecular modeling and simulation: an interdisciplinary guide: an interdisciplinary guide*. Springer Science & Business Media: 2010; Vol. 21.

33. Rauwerdink, A.; Kazlauskas, R. J., How the Same Core Catalytic Machinery Catalyzes 17 Different Reactions: the Serine-Histidine-Aspartate Catalytic Triad of α/β-Hydrolase Fold Enzymes. *ACS catalysis* **2015,** *5* (10), 6153-6176.

34. Michaelis L, M. M., Die Kinetik der Invertinwirkung. *Biochem Z* **1913,** *49*, 333–369.

35. Fischer, E., Einfluss der Configuration auf die Wirkung der Enzyme. *European Journal of Inorganic Chemistry* **1894,** *27* (3), 2985-2993.

36. Meierhenrich, U., *Amino Acids and the Asymmetry of Life: Caught in the Act of Formation*. Springer: 2008.

37. Moss, G. P., Basic terminology of stereochemistry (IUPAC Recommendations 1996). *Pure Appl. Chem.* **1996,** *68* (12), 2193-2222.

38. Straathof, A. J.; Adlercreutz, P., *Applied biocatalysis*. CRC Press: 2000.

39. Koshland, D. E., Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **1958,** *44* (2), 98-104.

40. Ma, B.; Kumar, S.; Tsai, C.-J.; Nussinov, R., Folding funnels and binding mechanisms. *Protein Engineering, Design and Selection* **1999,** *12* (9), 713-720.

41. Tsai, C.-J.; Kumar, S.; Ma, B.; Nussinov, R., Folding funnels, binding funnels, and protein function. *Protein Science* **1999,** *8* (6), 1181-1190.

42. Tobi, D.; Bahar, I., Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (52), 18908-18913.

43. Csermely, P.; Palotai, R.; Nussinov, R., Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences* **2010,** *35* (10), 539-546.

44. Garrett, B. C.; Truhlar, D. G., Variational transition state theory. In *Theory and Applications of Computational Chemistry*, Elsevier: 2005; pp 67-87.

45. Marti, S.; Roca, M.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J., Theoretical insights in enzyme catalysis. *Chemical Society reviews* **2004,** *33* (2), 98-107.

46. Lyne, P. D.; Mulholland, A. J.; Richards, W. G., Insights into chorismate mutase catalysis from a combined QM/MM simulation of the enzyme reaction. *Journal of the American Chemical Society* **1995,** *117* (45), 11345-11350.

47. Pauling, L., Molecular Architecture and Biological Reactions. *Chemical & Engineering News Archive* **1946,** *24* (10), 1375-1377.

48. Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G., How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science* **2004,** *303* (5655), 186.

49. Warshel, A.; Bora, R. P., Chapter 1 Perspective on Computer Modelling of Enzymatic Reactions. In *Simulating Enzyme Reactivity: Computational Methods in Enzyme Catalysis*, The Royal Society of Chemistry: 2017; pp 1-30.

50. Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H., Electrostatic basis for enzyme catalysis. *Chemical reviews* **2006,** *106* (8), 3210-35.

51. Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I., Enzymatic Effects on Reactant and Transition States. The Case of Chalcone Isomerase. *Journal of the American Chemical Society* **2007,** *129* (29), 9117-9124.

52. Szefczyk, B.; Claeyssens, F.; Mulholland, A. J.; Sokalski, W. A., Quantum chemical analysis of reaction paths in chorismate mutase: Conformational effects and electrostatic stabilization. *International Journal of Quantum Chemistry* **2007,** *107* (12), 2274-2285.

53. Villà, J.; Warshel, A., Energetics and Dynamics of Enzymatic Reactions. *The Journal of Physical Chemistry B* **2001,** *105* (33), 7887-7907.

54. Warshel, A., Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *Journal of Biological Chemistry* **1998,** *273* (42), 27035-27038.

55. Warshel, A., Energetics of enzyme catalysis. *Proceedings of the National Academy of Sciences* **1978,** *75* (11), 5250-5254.

56. Fried, S. D.; Bagchi, S.; Boxer, S. G., Extreme electric fields power catalysis in the active site of ketosteroid isomerase. *Science* **2014,** *346* (6216), 1510-4.

57. Fried, S. D.; Boxer, S. G., Electric Fields and Enzyme Catalysis. *Annual Review of Biochemistry* **2017,** *86* (1), 387-415.

58. Kamerlin, S. C.; Warshel, A., At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins* **2010,** *78* (6), 1339-75.

59. Warshel, A.; Bora, R. P., Perspective: Defining and quantifying the role of dynamics in enzyme catalysis. *The Journal of Chemical Physics* **2016,** *144* (18), 180901.

60. Schwartz, S. D.; Schramm, V. L., Enzymatic transition states and dynamic motion in barrier crossing. *Nature chemical biology* **2009,** *5* (8), 551.

61. Wang, Z.; Singh, P.; Czekster, C. M.; Kohen, A.; Schramm, V. L., Protein mass-modulated effects in the catalytic mechanism of dihydrofolate reductase: beyond

promoting vibrations. *Journal of the American Chemical Society* **2014,** *136* (23), 8333-8341.

62. Kohen, A., Role of dynamics in enzyme catalysis: substantial versus semantic controversies. *Accounts of chemical research* **2014,** *48* (2), 466-473.

63. Henzler-Wildman, K.; Kern, D., Dynamic personalities of proteins. *Nature* **2007,** *450*, 964.

64. Hay, S.; Scrutton, N. S., Good vibrations in enzyme-catalysed reactions. *Nature Chemistry* **2012,** *4*, 161.

65. Nagel, Z. D.; Klinman, J. P., A 21st century revisionist's view at a turning point in enzymology. *Nature chemical biology* **2009,** *5* (8), 543-50.

66. Caratzoulas, S.; Schwartz, S. D., A computational method to discover the existence of promoting vibrations for chemical reactions in condensed phases. *The Journal of Chemical Physics* **2001,** *114* (7), 2910-2918.

67. Antoniou, D.; Caratzoulas, S.; Kalyanaraman, C.; Mincer, J. S.; Schwartz, S. D., Barrier passage and protein dynamics in enzymatically catalyzed reactions. *European journal of biochemistry* **2002,** *269* (13), 3103-12.

68. Schramm, V. L.; Schwartz, S. D., Promoting Vibrations and the Function of Enzymes. Emerging Theoretical and Experimental Convergence. *Biochemistry* **2018**.

69. Benkovic, S. J.; Hammes-Schiffer, S., A perspective on enzyme catalysis. *Science* **2003,** *301* (5637), 1196-1202.

70. Bhabha, G.; Lee, J.; Ekiert, D. C.; Gam, J.; Wilson, I. A.; Dyson, H. J.; Benkovic, S. J.; Wright, P. E., A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* **2011,** *332* (6026), 234-238.

71. Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J., Taking Ockham's razor to enzyme dynamics and catalysis. *Nature chemistry* **2012,** *4* (3), 169.

72. Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H.-J., High-Accuracy Computation of Reaction Barriers in Enzymes. *Angewandte Chemie International Edition* **2006,** *45* (41), 6856-6859.

73. Luk, L. Y. P.; Javier Ruiz-Pernía, J.; Dawson, W. M.; Roca, M.; Loveridge, E. J.; Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J.; Tuñón, I.; Moliner, V.; Allemann, R. K., Unraveling the role of protein dynamics in dihydrofolate reductase catalysis. *Proceedings of the National Academy of Sciences* **2013,** *110* (41), 16344.

74. McGeagh, J. D.; Ranaghan, K. E.; Mulholland, A. J., Protein dynamics and enzyme catalysis: Insights from simulations. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2011,** *1814* (8), 1077-1092.

75. Es, I.; Vieira, J. D.; Amaral, A. C., Principles, techniques, and applications of biocatalyst immobilization for industrial application. *Appl Microbiol Biotechnol* **2015,** *99* (5), 2065-82.

76. Sheldon, R. A.; van Pelt, S., Enzyme immobilisation in biocatalysis: why, what and how. *Chemical Society reviews* **2013,** *42* (15), 6223-6235.

77. Zaks, A.; Klibanov, A. M., Enzymatic catalysis in organic media at 100 degrees C. *Science* **1984,** *224* (4654), 1249-51.

78. Itoh, T., Ionic Liquids as Tool to Improve Enzymatic Organic Synthesis. *Chemical reviews* **2017,** *117* (15), 10567-10607.

79. Patel, R. N., Biocatalysis for synthesis of pharmaceuticals. *Bioorganic & medicinal chemistry* **2017**.

80. Xu, Z.; Sha, Y.; Liu, C.; Li, S.; Liang, J.; Zhou, J.; Xu, H., L-Ribose isomerase and mannose-6-phosphate isomerase: properties and applications for L-ribose production. *Appl Microbiol Biotechnol* **2016,** *100* (21), 9003-9011.

81. Hara, R.; Hirai, K.; Suzuki, S.; Kino, K., A chemoenzymatic process for amide bond formation by an adenylating enzyme-mediated mechanism. *Scientific Reports* **2018,** *8* (1), 2950.

82. Ma, S. K.; Gruber, J.; Davis, C.; Newman, L.; Gray, D.; Wang, A.; Grate, J.; Huisman, G. W.; Sheldon, R. A., A green-by-design biocatalytic process for atorvastatin intermediate. *Green Chemistry* **2010,** *12* (1), 81-86.

83. Li, T.; Liang, J.; Ambrogelly, A.; Brennan, T.; Gloor, G.; Huisman, G.; Lalonde, J.; Lekhal, A.; Mijts, B.; Muley, S.; Newman, L.; Tobin, M.; Wong, G.; Zaks, A.; Zhang, X., Efficient, Chemoenzymatic Process for Manufacture of the Boceprevir Bicyclic [3.1.0]Proline Intermediate Based on Amine Oxidase-Catalyzed Desymmetrization. *Journal of the American Chemical Society* **2012,** *134* (14), 6467-6472.

84. Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J., Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **2010,** *329* (5989), 305-9.

85. Jose, C.; Toledo, M. V.; Briand, L. E., Enzymatic kinetic resolution of racemic ibuprofen: past, present and future. *Critical reviews in biotechnology* **2016,** *36* (5), 891-903.

86. Parmar, A.; Kumar, H.; Marwaha, S. S.; Kennedy, J. F., Advances in enzymatic transformation of penicillins to 6-aminopenicillanic acid (6-APA). *Biotechnology advances* **2000,** *18* (4), 289-301.

87. Mahmoudian, M.; Noble, D.; Drake, C. S.; Middleton, R. F.; Montgomery, D. S.; Piercey, J. E.; Ramlakhan, D.; Todd, M.; Dawson, M. J., An efficient process for production of n-acetylneuraminic acid using n-acetylneuraminic acid aldolase. *Enzyme and Microbial Technology* **1997,** *20* (5), 393-400.

88. Sheldon, R., *Organic synthesis. Past, present and future*. 1992; Vol. 23, p 903-906.

89. Anastas, P. T. W., J. C. , *Green Chemistry: Theory and Practice*. Oxford University Press: 1998.

90. Rozzell, J. D., Commercial scale biocatalysis: myths and realities. *Bioorganic & medicinal chemistry* **1999,** *7* (10), 2253-61.

91. Vigsoe, D.; Jürgensen, E.; Kvistgaard, M.; Wolf, O., *The assessment of future environmental and economic impacts of process-integrated biocatalysts*. European Commission, Joint Research Centre: 2002.

92. Clouthier, C. M.; Pelletier, J. N., Expanding the organic toolbox: a guide to integrating biocatalysis in synthesis. *Chemical Society reviews* **2012,** *41* (4), 1585-1605.

93. Ligon, B. L., Penicillin: its discovery and early development. *Semin Pediatr Infect Dis* **2004,** *15* (1), 52-7.

94. Fleming, A., On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. 1929. *Bull World Health Organ* **2001,** *79* (8), 780-90.

95. Rolinson, G. N.; Geddes, A. M., The 50th anniversary of the discovery of 6-aminopenicillanic acid (6-APA). *Int J Antimicrob Agents* **2007,** *29* (1), 3-8.

96. Sutherland, R.; Croydon, E. A.; Rolinson, G. N., Amoxycillin: a new semi-synthetic penicillin. *Br Med J* **1972,** *3* (5817), 13-6.

97. Lindsley, C. W., 2012 Trends and Statistics for Prescription Medications in the United States: CNS Therapeutics Continue to Hold Leading Positions. *ACS Chemical Neuroscience* **2013,** *4* (8), 1133-1135.

98. Kern, M.; Roiban, G.-D.; Fosberry, A.; Snajdrova, R., CHAPTER 9 GSK: Biocatalyst Discovery and Optimisation. In *Biocatalysis: An Industrial Perspective*, The Royal Society of Chemistry: 2018; pp 257-275.

99. Calcaterra, A.; D'Acquarica, I., The market of chiral drugs: Chiral switches versus de novo enantiomerically pure compounds. *Journal of Pharmaceutical and Biomedical Analysis* **2018,** *147*, 323-340.

100. Nunez, M. C.; Garcia-Rubino, M. E.; Conejo-Garcia, A.; Cruz-Lopez, O.; Kimatrai, M.; Gallo, M. A.; Espinosa, A.; Campos, J. M., Homochiral drugs: a demanding tendency of the pharmaceutical industry. *Curr Med Chem* **2009,** *16* (16), 2064-74.

101. McConathy, J.; Owens, M. J., Stereochemistry in Drug Action. *Prim Care Companion J Clin Psychiatry* **2003,** *5* (2), 70-73.

102. Caner, H.; Groner, E.; Levy, L.; Agranat, I., Trends in the development of chiral drugs. *Drug discovery today* **2004,** *9* (3), 105-10.

103. Agranat, I.; Wainschtein, S. R.; Zusman, E. Z., The predicated demise of racemic new molecular entities is an exaggeration. *Nature reviews. Drug discovery* **2012,** *11* (12), 972-3.

104. Drugs@FDA Glossary of Terms. https://www.fda.gov/drugs/informationondrugs/ucm079436.htm.

105. Lenz, W., A short history of thalidomide embryopathy. *Teratology* **1988,** *38* (3), 203-15.

106. McBride, W. G., Thalidomide and congenital abnormalities. *The Lancet* **1961,** *278* (7216), 1358.

107. Heger, W.; Schmahl, H. J.; Klug, S.; Felies, A.; Nau, H.; Merker, H. J.; Neubert, D., Embryotoxic effects of thalidomide derivatives in the non-human primate callithrix jacchus. IV. Teratogenicity of micrograms/kg doses of the EM12 enantiomers. *Teratog Carcinog Mutagen* **1994,** *14* (3), 115-22.

108. Reist, M.; Carrupt, P. A.; Francotte, E.; Testa, B., Chiral inversion and hydrolysis of thalidomide: mechanisms and catalysis by bases and serum albumin, and chiral stability of teratogenic metabolites. *Chem Res Toxicol* **1998,** *11* (12), 1521-8.

109. Krantz, J. C., Jr., The Kefauver-Harris amendment after sixteen years. *Military medicine* **1978,** *143* (12), 883.

110. Shah, R. R., Thalidomide, drug safety and early drug regulation in the UK. *Adverse Drug React Toxicol Rev* **2001,** *20* (4), 199-255.

111. Evans, A. M., Comparative pharmacology of S(+)-ibuprofen and (RS)-ibuprofen. *Clin Rheumatol* **2001,** *20 Suppl 1*, S9-14.

112. Rainsford, K. D., Fifty years since the discovery of ibuprofen. *Inflammopharmacology* **2011,** *19* (6), 293-7.

113. Agranat, I.; Caner, H.; Caldwell, J., Putting chirality to work: the strategy of chiral switches. *Nature reviews. Drug discovery* **2002,** *1* (10), 753-68.

114. Ward, T. J.; Ward, K. D., Chiral separations: a review of current topics and trends. *Anal Chem* **2012,** *84* (2), 626-35.

115. Rossi, D.; Tarantino, M.; Rossino, G.; Rui, M.; Juza, M.; Collina, S., Approaches for multi-gram scale isolation of enantiomers for drug discovery. *Expert opinion on drug discovery* **2017,** *12* (12), 1253-1269.

116. Albarrán-Velo, J.; González-Martínez, D.; Gotor-Fernández, V., *Stereoselective biocatalysis: A mature technology for the asymmetric synthesis of pharmaceutical building blocks*. 2017; Vol. 36, p 1-29.

117. Burke, D.; Henderson, D. J., Chirality: a blueprint for the future. *Br J Anaesth* **2002,** *88* (4), 563-76.

118. Noyori, R.; Ohkuma, T.; Kitamura, M.; Takaya, H.; Sayo, N.; Kumobayashi, H.; Akutagawa, S., Asymmetric hydrogenation of .beta.-keto carboxylic esters. A practical, purely chemical access to .beta.-hydroxy esters in high enantiomeric purity. *Journal of the American Chemical Society* **1987,** *109* (19), 5856-5858.

119. Gnas, Y.; Glorius, F., Chiral Auxiliaries - Principles and Recent Applications. *Synthesis* **2006,** *2006* (12), 1899-1930.

120. Ricci, A., Asymmetric Organocatalysis at the Service of Medicinal Chemistry. *ISRN Organic Chemistry* **2014,** *2014*, 29.

121. Carvalho, C. A.; Fonseca, D. T.; Mattos, C. M.; Oliveira, D. M.; Lemos, L. T.; Molinari, F.; Romano, D.; Serra, I., Recent Advances in Lipase-Mediated Preparation of Pharmaceuticals and Their Intermediates. *International Journal of Molecular Sciences* **2015,** *16* (12).

122. Gotor-Fernández, V.; Brieva, R.; Gotor, V., Lipases: Useful biocatalysts for the preparation of pharmaceuticals. *Journal of Molecular Catalysis B: Enzymatic* **2006,** *40* (3), 111-120.

123. Sikora, A.; Siódmiak, T.; Marszałł, M. P., Kinetic Resolution of Profens by EnantioselectiveEsterification Catalyzed by Candida antarctica and Candida rugosa Lipases. *Chirality* **2014,** *26* (10), 663-669.

124. Habibi, Z.; Mohammadi, M.; Yousefi, M., Enzymatic hydrolysis of racemic ibuprofen esters using Rhizomucor miehei lipase immobilized on different supports. *Process Biochemistry* **2013,** *48* (4), 669-676.

125. Long, W. S.; Kow, P. C.; Kamaruddin, A. H.; Bhatia, S., Comparison of kinetic resolution between two racemic ibuprofen esters in an enzymic membrane reactor. *Process Biochemistry* **2005,** *40* (7), 2417-2425.

126. Tomasz, S.; Jan, K. R.; Michal, P. M., Application of Lipases from Candida rugosa in the Enantioselective Esterification of (R,S)-Ibuprofen. *Current Organic Chemistry* **2012,** *16* (8), 972-977.

127. Desai, S. B.; Argade, N. P.; Ganesh, K. N., Remarkable Chemo-, Regio-, and Enantioselectivity in Lipase-Catalyzed Hydrolysis: Efficient Resolution of (+/-)-threo-Ethyl 3-(4-Methoxyphenyl)-2,3-diacetoxypropionate Leading to Chiral Intermediates of (+)-Diltiazem. *The Journal of organic chemistry* **1996,** *61* (19), 6730-6732.

128. Zhao, L.-L.; Pan, J.; Xu, J.-H., Efficient production of diltiazem chiral intermediate using immobilized lipase from Serratia marcescens. *Biotechnology and Bioprocess Engineering* **2010,** *15* (2), 199-207.

129. Liang, J.; Lalonde, J.; Borup, B.; Mitchell, V.; Mundorff, E.; Trinh, N.; Kochrekar, D. A.; Nair Cherat, R.; Pai, G. G., Development of a biocatalytic process as an alternative to the (−)-DIP-Cl-mediated asymmetric reduction of a key intermediate of montelukast. *Organic Process Research & Development* **2009,** *14* (1), 193-198.

130. Ghislieri, D.; Turner, N. J., Biocatalytic Approaches to the Synthesis of Enantiomerically Pure Chiral Amines. *Topics in Catalysis* **2014,** *57* (5), 284-300.

131. Ismail, H.; Lau Rute, M.; van Rantwijk, F.; Sheldon Roger , A., Fully Enzymatic Resolution of Chiral Amines: Acylation and Deacylation in the Presence of Candida antarctica Lipase B. *Advanced Synthesis & Catalysis* **2008,** *350* (10), 1511-1516.

132. Fuchs, M.; Farnberger, J. E.; Kroutil, W., The Industrial Age of Biocatalytic Transamination. *European J Org Chem* **2015,** *2015* (32), 6965-6982.

133. Kim, D.; Wang, L.; Beconi, M.; Eiermann, G. J.; Fisher, M. H.; He, H.; Hickey, G. J.; Kowalchick, J. E.; Leiting, B.; Lyons, K.; Marsilio, F.; McCann, M. E.; Patel, R. A.; Petrov, A.; Scapin, G.; Patel, S. B.; Roy, R. S.; Wu, J. K.; Wyvratt, M. J.; Zhang, B. B.; Zhu, L.; Thornberry, N. A.; Weber, A. E., (2R)-4-oxo-4-[3-(trifluoromethyl)-5,6-dihydro[1,2,4]triazolo[4,3-a]pyrazin-7(8H)- yl]-1-(2,4,5-trifluorophenyl)butan-2-amine: a potent, orally active dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. *Journal of medicinal chemistry* **2005,** *48* (1), 141-51.

134. Hansen, K. B.; Hsiao, Y.; Xu, F.; Rivera, N.; Clausen, A.; Kubryk, M.; Krska, S.; Rosner, T.; Simmons, B.; Balsells, J.; Ikemoto, N.; Sun, Y.; Spindler, F.; Malan, C.; Grabowski, E. J.; Armstrong, J. D., 3rd, Highly efficient asymmetric synthesis of sitagliptin. *J Am Chem Soc* **2009,** *131* (25), 8798-804.

135. Chung, C. K.; Bulger, P. G.; Kosjek, B.; Belyk, K. M.; Rivera, N.; Scott, M. E.; Humphrey, G. R.; Limanto, J.; Bachert, D. C.; Emerson, K. M., Process Development of C–N Cross-Coupling and Enantioselective Biocatalytic Reactions for the Asymmetric Synthesis of Niraparib. *Organic Process Research & Development* **2014,** *18* (1), 215-227.

136. Girardin, M.; Ouellet, S. G.; Gauvreau, D.; Moore, J. C.; Hughes, G.; Devine, P. N.; O'Shea, P. D.; Campeau, L.-C., Convergent Kilogram-Scale Synthesis of Dual Orexin Receptor Antagonist. *Organic Process Research & Development* **2013,** *17* (1), 61-68.

137. Frodsham, L.; Golden, M.; Hard, S.; Kenworthy, M. N.; Klauber, D. J.; Leslie, K.; Macleod, C.; Meadows, R. E.; Mulholland, K. R.; Reilly, J.; Squire, C.; Tomasi, S.; Watt, D.; Wells, A. S., Use of ω-Transaminase Enzyme Chemistry in the Synthesis of a JAK2 Kinase Inhibitor. *Organic Process Research & Development* **2013,** *17* (9), 1123-1130.

138. Grogan, G., Synthesis of chiral amines using redox biocatalysis. *Current Opinion in Chemical Biology* **2018,** *43*, 15-22.

139. Bui, S.; Steiner, R. A., New insight into cofactor-free oxygenation from combined experimental and computational approaches. *Current opinion in structural biology* **2016,** *41* (Supplement C), 109-118.

140. Ghislieri, D.; Green, A. P.; Pontini, M.; Willies, S. C.; Rowles, I.; Frank, A.; Grogan, G.; Turner, N. J., Engineering an Enantioselective Amine Oxidase for the Synthesis of Pharmaceutical Building Blocks and Alkaloid Natural Products. *Journal of the American Chemical Society* **2013,** *135* (29), 10863-10869.

141. Mangas-Sanchez, J.; France, S. P.; Montgomery, S. L.; Aleku, G. A.; Man, H.; Sharma, M.; Ramsden, J. I.; Grogan, G.; Turner, N. J., Imine reductases (IREDs). *Current Opinion in Chemical Biology* **2017,** *37*, 19-25.

142. Mitsukura, K.; Kuramoto, T.; Yoshida, T.; Kimoto, N.; Yamamoto, H.; Nagasawa, T., A NADPH-dependent (S)-imine reductase (SIR) from Streptomyces sp. GF3546 for asymmetric synthesis of optically active amines: purification, characterization, gene cloning, and expression. *Applied Microbiology and Biotechnology* **2013,** *97* (18), 8079-8086.

143. Mitsukura, K.; Suzuki, M.; Shinoda, S.; Kuramoto, T.; Yoshida, T.; Nagasawa, T., Purification and Characterization of a Novel (R)-Imine Reductase from Streptomyces sp. GF3587. *Bioscience, Biotechnology, and Biochemistry* **2011,** *75* (9), 1778-1782.

144. Roiban, G. D.; Kern, M.; Liu, Z.; Hyslop, J.; Tey Pei, L.; Levine Matthew, S.; Jordan Lydia, S.; Brown Kristin, K.; Hadi, T.; Ihnken Leigh Anne, F.; Brown Murray, J. B., Efficient Biocatalytic Reductive Aminations by Extending the Imine Reductase Toolbox. *ChemCatChem* **2017,** *9* (24), 4475-4479.

145. Schrittwieser, J. H.; Velikogne, S.; Hall, M.; Kroutil, W., Artificial Biocatalytic Linear Cascades for Preparation of Organic Molecules. *Chemical reviews* **2018,** *118* (1), 270-348.

146. Heath Rachel, S.; Pontini, M.; Hussain, S.; Turner Nicholas, J., Combined Imine Reductase and Amine Oxidase Catalyzed Deracemization of Nitrogen Heterocycles. *ChemCatChem* **2015,** *8* (1), 117-120.

147. Gomm, A.; O'Reilly, E., Transaminases for chiral amine synthesis. *Current Opinion in Chemical Biology* **2018,** *43*, 106-112.

148. Zhang, K.; Li, H.; Cho, K. M.; Liao, J. C., Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** *107* (14), 6234-9.

149. Sutton, P. W.; Adams, J. P.; Archer, I.; Auriol, D.; Avi, M.; Branneby, C.; Collis, A. J.; Dumas, B.; Eckrich, T.; Fotheringham, I.; ter Halle, R.; Hanlon, S.; Hansen, M.; Holt-Tiffin, K. E.; Howard, R. M.; Huisman, G. W.; Iding, H.; Kiewel, K.; Kittelmann, M.; Kupfer, E.; Laumen, K.; Lefèvre, F.; Luetz, S.; Mangan, D. P.; Martin, V. A.; Meyer, H.-P.; Moody, T. S.; Osorio-Lozada, A.; Robins, K.; Snajdrova, R.; Truppo, M. D.; Wells, A.; Wirz, B.; Wong, J. W., Biocatalysis in the Fine Chemical and Pharmaceutical Industries. In *Practical Methods for Biocatalysis and Biotransformations 2*, John Wiley & Sons, Ltd: 2012; pp 1-59.

150. Choi, J. M.; Han, S. S.; Kim, H. S., Industrial applications of enzyme biocatalysis: Current status and future aspects. *Biotechnology advances* **2015,** *33* (7), 1443-54.

151. Sheldon, R. A.; Pereira, P. C., Biocatalysis engineering: the big picture. *Chemical Society reviews* **2017,** *46* (10), 2678-2691.

152. Bommarius, A. S.; Blum, J. K.; Abrahamson, M. J., Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst. *Current Opinion in Chemical Biology* **2011,** *15* (2), 194-200.

153. Huang, P.-S.; Boyken, S. E.; Baker, D., The coming of age of de novo protein design. *Nature* **2016,** *537*, 320.

154. Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D., Kemp elimination catalysts by computational enzyme design. *Nature* **2008,** *453* (7192), 190-5.

155. Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D., De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008,** *319* (5868), 1387.

156. Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St.Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D., Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **2010,** *329* (5989), 309.

157. Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Röthlisberger, D.; Baker, D., New algorithms and an in silico benchmark for computational enzyme design. *Protein Science* **2006,** *15* (12), 2785-2794.

158. Kiss, G.; Çelebi‐Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N., Computational Enzyme Design. *Angewandte Chemie International Edition* **2013,** *52* (22), 5700-5725.

159. Khersonsky, O.; Kiss, G.; Röthlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S., Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proceedings of the National Academy of Sciences* **2012,** *109* (26), 10358.

160. Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R.; Grutter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D., Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **2013,** *503* (7476), 418-21.

161. Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K., Engineering the third wave of biocatalysis. *Nature* **2012,** *485* (7397), 185-94.

162. Jaeger, K.-E.; Eggert, T., Enantioselective biocatalysis optimized by directed evolution. *Current Opinion in Biotechnology* **2004,** *15* (4), 305-313.

163. Leung, D. W., A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1989,** *1*, 11-15.

164. Stemmer, W. P. C., Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **1994,** *370*, 389.

165. Leemhuis, H.; Kelly, R. M.; Dijkhuizen, L., Directed evolution of enzymes: Library screening strategies. *IUBMB Life* **2009,** *61* (3), 222-228.

166. Martis, E.; R, R.; R.R, B., *High-Throughput Screening: The Hits and Leads of Drug Discovery- An Overview*. 2011; Vol. 1, p 02-10.

167. Varadarajan, N.; Cantor, J. R.; Georgiou, G.; Iverson, B. L., Construction and flow cytometric screening of targeted enzyme libraries. *Nature protocols* **2009,** *4*, 893.

168. Yang, G.; Withers Stephen, G., Ultrahigh‐Throughput FACS‐Based Screening for Directed Enzyme Evolution. *Chembiochem : a European journal of chemical biology* **2009,** *10* (17), 2704-2715.

169. Obexer, R.; Pott, M.; Zeymer, C.; Griffiths, A. D.; Hilvert, D., Efficient laboratory evolution of computationally designed enzymes with low starting activities using fluorescence-activated droplet sorting. *Protein Engineering, Design and Selection* **2016,** *29* (9), 355-366.

170. Obexer, R.; Godina, A.; Garrabou, X.; Mittl, P. R. E.; Baker, D.; Griffiths, A. D.; Hilvert, D., Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nature Chemistry* **2016,** *9*, 50.

171. Goldsmith, M.; Tawfik, D. S., Enzyme engineering: reaching the maximal catalytic efficiency peak. *Current opinion in structural biology* **2017,** *47*, 140-150.

172. Romero, P. A.; Tran, T. M.; Abate, A. R., Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proceedings of the National Academy of Sciences* **2015,** *112* (23), 7159.

173. Fowler, D. M.; Fields, S., Deep mutational scanning: a new style of protein science. *Nature methods* **2014,** *11*, 801.

174. Wrenbeck, E. E.; Azouz, L. R.; Whitehead, T. A., Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nature Communications* **2017,** *8*, 15695.

175. Starr Tyler, N.; Thornton Joseph, W., Epistasis in protein evolution. *Protein Science* **2016,** *25* (7), 1204-1218.

176. Dalby, P. A., Strategy and success for the directed evolution of enzymes. *Current opinion in structural biology* **2011,** *21* (4), 473-80.

177. Turner, N. J., Directed evolution drives the next generation of biocatalysts. *Nature chemical biology* **2009,** *5* (8), 567-73.

178. Gupta, R. D.; Tawfik, D. S., Directed enzyme evolution via small and effective neutral drift libraries. *Nature methods* **2008,** *5* (11), 939-42.

179. Reetz, M. T., Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future. *Journal of the American Chemical Society* **2013,** *135* (34), 12480-12496.

180. Reetz Manfred, T., Laboratory Evolution of Stereoselective Enzymes: A Prolific Source of Catalysts for Asymmetric Reactions. *Angewandte Chemie International Edition* **2010,** *50* (1), 138-174.

181. Reetz, M. T.; Carballeira, J. D., Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature protocols* **2007,** *2* (4), 891-903.

182. Acevedo-Rocha, C. G.; Hoebenreich, S.; Reetz, M. T., Iterative Saturation Mutagenesis: A Powerful Approach to Engineer Proteins by Systematically Simulating

Darwinian Evolution. In *Directed Evolution Library Creation: Methods and Protocols*, Gillam, E. M. J.; Copp, J. N.; Ackerley, D., Eds. Springer New York: New York, NY, 2014; pp 103-128.

183. Reetz, M. T.; Bocola, M.; Carballeira, J. D.; Zha, D.; Vogel, A., Expanding the Range of Substrate Acceptance of Enzymes: Combinatorial Active-Site Saturation Test. *Angewandte Chemie International Edition* **2005,** *44* (27), 4192-4196.

184. Reetz, M. T.; Soni, P.; Fernandez, L.; Gumulya, Y.; Carballeira, J. D., Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method. *Chemical Communications* **2010,** *46* (45), 8657-8658.

185. Reetz, M. T.; Prasad, S.; Carballeira, J. D.; Gumulya, Y.; Bocola, M., Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *J Am Chem Soc* **2010,** *132* (26), 9144-52.

186. Sun, Z.; Lonsdale, R.; Kong, X.-D.; Xu, J.-H.; Zhou, J.; Reetz, M. T., Reshaping an Enzyme Binding Pocket for Enhanced and Inverted Stereoselectivity: Use of Smallest Amino Acid Alphabets in Directed Evolution. *Angewandte Chemie International Edition* **2015,** *54* (42), 12410-12415.

187. Sun, Z.; Lonsdale, R.; Wu, L.; Li, G.; Li, A.; Wang, J.; Zhou, J.; Reetz, M. T., Structure-guided triple-code saturation mutagenesis: efficient tuning of the stereoselectivity of an epoxide hydrolase. *ACS Catalysis* **2016,** *6* (3), 1590-1597.

188. Damborsky, J.; Brezovsky, J., Computational tools for designing and engineering biocatalysts. *Curr Opin Chem Biol* **2009,** *13* (1), 26-34.

189. Braiuca, P.; Ebert, C.; Basso, A.; Linda, P.; Gardossi, L., Computational methods to rationalize experimental strategies in biocatalysis. *Trends in biotechnology* **2006,** *24* (9), 419-25.

190. Pavelka, A.; Chovancova, E.; Damborsky, J., HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic acids research* **2009,** *37* (Web Server issue), W376-83.

191. Fox, R. J.; Davis, S. C.; Mundorff, E. C.; Newman, L. M.; Gavrilovic, V.; Ma, S. K.; Chung, L. M.; Ching, C.; Tam, S.; Muley, S.; Grate, J.; Gruber, J.; Whitman, J. C.; Sheldon, R. A.; Huisman, G. W., Improving catalytic function by ProSAR-driven enzyme evolution. *Nature Biotechnology* **2007,** *25*, 338.

192. Silberg, J. J.; Endelman, J. B.; Arnold, F. H., SCHEMA-Guided Protein Recombination. In *Methods in enzymology*, Academic Press: 2004; Vol. 388, pp 35-42.

193. Kuipers, R. K.; Joosten, H. J.; van Berkel, W. J.; Leferink, N. G.; Rooijen, E.; Ittmann, E.; van Zimmeren, F.; Jochens, H.; Bornscheuer, U.; Vriend, G.; dos Santos, V. A.; Schaap, P. J., 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins* **2010,** *78* (9), 2101-13.

194. Kourist, R.; Jochens, H.; Bartsch, S.; Kuipers, R.; Padhi, S. K.; Gall, M.; Bottcher, D.; Joosten, H. J.; Bornscheuer, U. T., The alpha/beta-hydrolase fold 3DM database (ABHDB) as a tool for protein engineering. *Chembiochem : a European journal of chemical biology* **2010,** *11* (12), 1635-43.

195. Suplatov, D.; Kirilin, E.; Takhaveev, V.; Svedas, V., Zebra: a web server for bioinformatic analysis of diverse protein families. *Journal of biomolecular structure & dynamics* **2014,** *32* (11), 1752-8.

196. Suplatov, D. A.; Besenmatter, W.; Švedas, V. K.; Svendsen, A., Bioinformatic analysis of alpha/beta-hydrolase fold enzymes reveals subfamily-specific positions responsible for discrimination of amidase and lipase activities†. *Protein Engineering, Design and Selection* **2012,** *25* (11), 689-697.

197. Monza, E.; Acebes, S.; Lucas, M. F.; Guallar, V., Molecular Modeling in Enzyme Design, Toward In Silico Guided Directed Evolution. In *Directed Enzyme Evolution: Advances and Applications*, Alcalde, M., Ed. Springer International Publishing: Cham, 2017; pp 257-284.

198. Korkegian, A.; Black, M. E.; Baker, D.; Stoddard, B. L., Computational Thermostabilization of an Enzyme. *Science* **2005,** *308* (5723), 857.

199. Kellogg Elizabeth, H.; Leaver‐Fay, A.; Baker, D., Role of conformational sampling in computing mutation‐induced changes in protein structure and stability. *Proteins: Structure, Function, and Bioinformatics* **2011,** *79* (3), 830-838.

200. Guerois, R.; Nielsen, J. E.; Serrano, L., Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. *Journal of molecular biology* **2002,** *320* (2), 369-387.

201. Buß, O.; Rudat, J.; Ochsenreither, K., FoldX as Protein Engineering Tool: Better Than Random Based Approaches? *Computational and Structural Biotechnology Journal* **2018,** *16*, 25-33.

202. Buß, O.; Muller, D.; Jager, S.; Rudat, J.; Rabe, K. S., Improvement in the Thermostability of a β‐Amino Acid Converting ω‐Transaminase by Using FoldX. *Chembiochem : a European journal of chemical biology* **2018,** *19* (4), 379-387.

203. Wijma, H. J.; Floor, R. J.; Jekel, P. A.; Baker, D.; Marrink, S. J.; Janssen, D. B., Computationally designed libraries for rapid enzyme stabilization. *Protein Engineering, Design and Selection* **2014,** *27* (2), 49-58.

204. Wijma, H. J.; Furst, M.; Janssen, D. B., A Computational Library Design Protocol for Rapid Improvement of Protein Stability: FRESCO. *Methods in molecular biology (Clifton, N.J.)* **2018,** *1685*, 69-85.

205. Schueler-Furman, O.; Wang, C.; Bradley, P.; Misura, K.; Baker, D., Progress in modeling of protein structures and interactions. *Science* **2005,** *310* (5748), 638-42.

206. Wijma Hein, J.; Floor Robert, J.; Bjelic, S.; Marrink Siewert, J.; Baker, D.; Janssen Dick, B., Enantioselective Enzymes by Computational Design and In Silico Screening. *Angewandte Chemie* **2015,** *127* (12), 3797-3801.

207. Gainza, P.; Roberts, K. E.; Georgiev, I.; Lilien, R. H.; Keedy, D. A.; Chen, C. Y.; Reza, F.; Anderson, A. C.; Richardson, D. C.; Richardson, J. S.; Donald, B. R., OSPREY: protein design with ensembles, flexibility, and provable algorithms. *Methods in enzymology* **2013,** *523*, 87-107.

208. Huang, P.-S.; Ban, Y.-E. A.; Richter, F.; Andre, I.; Vernon, R.; Schief, W. R.; Baker, D., RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design. *PLoS ONE* **2011,** *6* (8), e24109.

209. Hediger, M. R.; De Vico, L.; Svendsen, A.; Besenmatter, W.; Jensen, J. H., A computational methodology to screen activities of enzyme variants. *PLoS One* **2012,** *7* (12), e49849.

210. Funke, S. A.; Otte, N.; Eggert, T.; Bocola, M.; Jaeger, K. E.; Thiel, W., Combination of computational prescreening and experimental library construction can

accelerate enzyme optimization by directed evolution. *Protein engineering, design & selection : PEDS* **2005,** *18* (11), 509-14.

211. Zheng, F.; Yang, W.; Ko, M.-C.; Liu, J.; Cho, H.; Gao, D.; Tong, M.; Tai, H.-H.; Woods, J. H.; Zhan, C.-G., Most Efficient Cocaine Hydrolase Designed by Virtual Screening of Transition States. *Journal of the American Chemical Society* **2008,** *130* (36), 12148-12155.

212. Gu, J.; Liu, M.; Guo, F.; Xie, W.; Lu, W.; Ye, L.; Chen, Z.; Yuan, S.; Yu, H., Virtual screening of mandelate racemase mutants with enhanced activity based on binding energy in the transition state. *Enzyme and Microbial Technology* **2014,** *55*, 121-127.

213. Grisewood, M. J.; Gifford, N. P.; Pantazes, R. J.; Li, Y.; Cirino, P. C.; Janik, M. J.; Maranas, C. D., OptZyme: computational enzyme redesign using transition state analogues. *PLoS One* **2013,** *8* (10), e75358.

214. Hediger, M. R.; De Vico, L.; Rannes, J. B.; Jäckel, C.; Besenmatter, W.; Svendsen, A.; Jensen, J. H., In silico screening of 393 mutants facilitates enzyme engineering of amidase activity in CalB. *PeerJ* **2013,** *1*, e145.

215. Sun, H.; Zhang, H.; Ang, E. L.; Zhao, H., Biocatalysis for the synthesis of pharmaceuticals and pharmaceutical intermediates. *Bioorganic & medicinal chemistry* **2017**.

216. Lalonde, J., Highly engineered biocatalysts for efficient small molecule pharmaceutical synthesis. *Current Opinion in Biotechnology* **2016,** *42*, 152-158.

217. Dorr, B. M.; Fuerst, D. E., Enzymatic amidation for industrial applications. *Current Opinion in Chemical Biology* **2018,** *43*, 127-133.

218. Strynadka, N. C.; James, M. N., Lysozyme: a model enzyme in protein crystallography. *Exs* **1996,** *75*, 185-222.

219. Warshel, A.; Levitt, M., Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of molecular biology* **1976,** *103* (2), 227-49.

220. Schrödinger, E., An undulatory theory of the mechanics of atoms and molecules. *Physical review* **1926,** *28* (6), 1049.

221. Hehre, W. J., *A Brief Guide to Molecular Mechanics and Quantum Chemical Calculations*. Wavefunction: 1998.

222. Hartree, D. R., The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Mathematical Proceedings of the Cambridge Philosophical Society* **1928,** *24* (1), 89-110.

223. Born, M. a. O., R, Zur Quantentheorie der Molekeln. *Ann. Phys.* **1927,** *389* (20), 457–484.

224. Lennard-Jones, J. E., The electronic structure of some diatomic molecules. *Transactions of the Faraday Society* **1929,** *25* (0), 668-686.

225. Jensen, F., *Introduction to Computational Chemistry*. 2 ed.; Wiley: 2007.

226. He, X.; Merz, K. M., Jr., Divide-and-Conquer Hartree-Fock Calculations on Proteins. *J Chem Theory Comput* **2010,** *6* (2), 405-411.

227. Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas. *Physical Review* **1964,** *136* (3B), B864-B871.

228. Pople, J. A.; Head‐Gordon, M.; Raghavachari, K., Quadratic configuration interaction. A general technique for determining electron correlation energies. *The Journal of Chemical Physics* **1987,** *87* (10), 5968-5975.

229. Paldus, J.; Čížek, J.; Shavitt, I., Correlation Problems in Atomic and Molecular Systems. IV. Extended Coupled-Pair Many-Electron Theory and Its Application to the B H 3 Molecule. *Physical Review A* **1972,** *5* (1), 50.

230. Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M., A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters* **1989,** *157* (6), 479-483.

231. Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965,** *140* (4A), A1133-A1138.

232. Marques, M. A. L.; Oliveira, M. J. T.; Burnus, T., Libxc: A library of exchange and correlation functionals for density functional theory. *Computer Physics Communications* **2012,** *183* (10), 2272-2281.

233. Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A* **1988,** *38* (6), 3098.

234. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review. B, Condensed matter* **1988,** *37* (2), 785-789.

235. Zhao, Y.; Schultz, N. E.; Truhlar, D., Exchange-correlation functional with broad accuracy for metallic and nonmetallic compounds, kinetics, and noncovalent interactions. AIP: 2005.

236. Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts* **2008,** *120* (1), 215-241.

237. Kristyán, S.; Pulay, P., Can (semi) local density functional theory account for the London dispersion forces? *Chemical physics letters* **1994,** *229* (3), 175-180.

238. Černý, J.; Hobza, P., Non-covalent interactions in biomacromolecules. *Physical Chemistry Chemical Physics* **2007,** *9* (39), 5291-5303.

239. Grimme, S., Density functional theory with London dispersion corrections. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011,** *1* (2), 211-228.

240. Grimme, S., Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J Comput Chem* **2004,** *25* (12), 1463-73.

241. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* **2010,** *132* (15), 154104.

242. Lonsdale, R.; Harvey, J. N.; Mulholland, A. J., Effects of Dispersion in Density Functional Based Quantum Mechanical/Molecular Mechanical Calculations on Cytochrome P450 Catalyzed Reactions. *J Chem Theory Comput* **2012,** *8* (11), 4637-45.

243. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc* **1996,** *118* (45), 11225-11236.

244. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J Comput Chem* **2004,** *25* (9), 1157-74.

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

246. Leach, A. R., *Molecular modelling: principles and applications*. Pearson education: 2001.

247. Mulholland, A. J., Chemical accuracy in QM/MM calculations on enzyme-catalysed reactions. *Chemistry Central Journal* **2007,** *1* (1), 19.

248. Sousa Sérgio, F.; Ribeiro António, J. M.; Neves Rui, P. P.; Brás Natércia, F.; Cerqueira Nuno, M. F. S. A.; Fernandes Pedro, A.; Ramos Maria, J., Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2016,** *7* (2), e1281.

249. Ryde, U.; Soderhjelm, P., Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods. *Chemical reviews* **2016,** *116* (9), 5520-66.

250. Liu, L.; Cui, G.; Fang, W. H., Excited States and Photochemistry of Chromophores in the Photoactive Proteins Explored by the Combined Quantum Mechanical and Molecular Mechanical Calculations. *Advances in protein chemistry and structural biology* **2015,** *100*, 255-84.

251. Senn, H. M.; Thiel, W., QM/MM Methods for Biological Systems. In *Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations*, Reiher, M., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; pp 173-290.

252. QSite, v., Schrödinger, LLC, New York, NY, 2014.

253. Murphy, R. B.; Philipp, D. M.; Friesner, R. A., A mixed quantum mechanics/molecular mechanics (QM/MM) method for large‐scale modeling of chemistry in protein environments. *Journal of Computational Chemistry* **2000,** *21* (16), 1442-1457.

254. Senn, H. M.; Thiel, W., QM/MM methods for biomolecular systems. *Angewandte Chemie (International ed. in English)* **2009,** *48* (7), 1198-229.

255. Field, M. J.; Bash, P. A.; Karplus, M., A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry* **1990,** *11* (6), 700-733.

256. Antila, H. S.; Salonen, E., Polarizable force fields. In *Biomolecular Simulations*, Springer: 2013; pp 215-241.

257. Singh, U. C.; Kollman, P. A., A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH3Cl + Cl− exchange reaction and gas phase protonation of polyethers. *Journal of Computational Chemistry* **1986,** *7* (6), 718-730.

258. Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J., A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives1Dedicated to Professor Keiji Morokuma in celebration of his 65th birthday.1. *Journal of Molecular Structure: THEOCHEM* **1999,** *461-462*, 1-21.

259. Swart, M., AddRemove: A new link model for use in QM/MM studies. *International Journal of Quantum Chemistry* **2002,** *91* (2), 177-183.

260. Lin, H.; Truhlar, D. G., QM/MM: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts* **2006,** *117* (2), 185.

261. Théry, V.; Rinaldi, D.; Rivail, J. L.; Maigret, B.; Ferenczy, G. G., Quantum mechanical computations on very large molecular systems: The local self‐consistent field method. *Journal of computational chemistry* **1994,** *15* (3), 269-282.

262. Gao, J.; Amara, P.; Alhambra, C.; Field, M. J., A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *The Journal of Physical Chemistry A* **1998,** *102* (24), 4714-4721.

263. Martí, S.; Moliner, V.; Tuñón, I., Improving the QM/MM description of chemical processes: a dual level strategy to explore the potential energy surface in very large systems. *Journal of chemical theory and computation* **2005,** *1* (5), 1008-1016.

264. Kästner, J.; Thiel, S.; Senn, H. M.; Sherwood, P.; Thiel, W., Exploiting QM/MM Capabilities in Geometry Optimization: A Microiterative Approach Using Electrostatic Embedding. *Journal of Chemical Theory and Computation* **2007,** *3* (3), 1064-1072.

265. Nash, S. G., A survey of truncated-Newton methods. *Journal of Computational and Applied Mathematics* **2000,** *124* (1), 45-59.

266. Bjelic, S.; Aqvist, J., Computational prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site. *Biochemistry* **2004,** *43* (46), 14521-8.

267. Labute, P., Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* **2009,** *75* (1), 187-205.

268. Chemical Computing Group Inc., S. S. W., Suite #910, Montreal, QC, Canada, H3A 2R7, Molecular Operating Environment (MOE), 2013.08. **2016**.

269. Lonsdale, R.; Harvey, J. N.; Mulholland, A. J., A practical guide to modelling enzyme-catalysed reactions. *Chemical Society reviews* **2012,** *41* (8), 3025-3038.

270. Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D. In *Scalable algorithms for molecular dynamics simulations on commodity clusters*, Proceedings of the 2006 ACM/IEEE conference on Supercomputing, ACM: 2006; p 84.

271. Desmond Molecular Dynamics System, v., D. E. Shaw Research, New York, NY, 2014. Maestro-Desmond Interoperability Tools, version 3.9, Schrödinger, New York, NY, 2014.

272. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926-935.

273. Shaw, K. E.; Woods, C. J.; Mulholland, A. J., Compatibility of quantum chemical methods and empirical (MM) water models in quantum mechanics/molecular mechanics liquid water simulations. *The Journal of Physical Chemistry Letters* **2009,** *1* (1), 219-223.

274. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* **1994,** *101* (5), 4177-4189.

275. Martyna, G. J.; Klein, M. L.; Tuckerman, M., Nosé–Hoover chains: the canonical ensemble via continuous dynamics. *The Journal of chemical physics* **1992,** *97* (4), 2635-2643.

276. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995,** *103* (19), 8577-8593.

277. Becke, A. D., A new mixing of Hartree‐Fock and local density‐functional theories. *The Journal of Chemical Physics* **1993,** *98* (2), 1372-1377.

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

279. Hariharan, P. C.; Pople, J. A., The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica chimica acta* **1973,** *28* (3), 213-222.

280. Kollar, J.; Frecer, V., How accurate is the description of ligand-protein interactions by a hybrid QM/MM approach? *Journal of molecular modeling* **2017,** *24* (1), 11.

281. Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J., Protocol for Computational Enzymatic Reactivity Based on Geometry Optimisation. *Chemphyschem : a European journal of chemical physics and physical chemistry* **2017,** *19* (6), 669-689.

282. Ghysels, A.; Woodcock Iii, H. L.; Larkin, J. D.; Miller, B. T.; Shao, Y.; Kong, J.; Neck, D. V.; Speybroeck, V. V.; Waroquier, M.; Brooks, B. R., Efficient calculation of QM/MM frequencies with the Mobile Block Hessian. *Journal of chemical theory and computation* **2011,** *7* (2), 496-514.

283. Herrmann, C.; Neugebauer, J.; Reiher, M., QM/MM vibrational mode tracking. *Journal of computational chemistry* **2008,** *29* (14), 2460-2470.

284. Ranaghan, K. E.; Mulholland, A. J., Investigations of enzyme-catalysed reactions with combined quantum mechanics/molecular mechanics (QM/MM) methods. *International Reviews in Physical Chemistry* **2010,** *29* (1), 65-133.

285. Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M., Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry* **1991,** *30* (24), 5826-5832.

286. Penning, T. M., The aldo-keto reductases (AKRs): Overview. *Chemico-biological interactions* **2015,** *234*, 236-46.

287. Ramana, K. V.; Srivastava, S. K., Aldose reductase: a novel therapeutic target for inflammatory pathologies. *The international journal of biochemistry & cell biology* **2010,** *42* (1), 17-20.

288. Petrash, J. M., All in the family: aldose reductase and closely related aldo-keto reductases. *Cellular and molecular life sciences : CMLS* **2004,** *61* (7-8), 737-49.

289. Maccari, R.; Ottana, R., Targeting aldose reductase for the treatment of diabetes complications and inflammatory diseases: new insights and future directions. *Journal of medicinal chemistry* **2015,** *58* (5), 2047-67.

290. Chatzopoulou, M.; Alexiou, P.; Kotsampasakou, E.; Demopoulos, V. J., Novel aldose reductase inhibitors: a patent survey (2006--present). *Expert opinion on therapeutic patents* **2012,** *22* (11), 1303-23.

291. Wilson, D. K.; Bohren, K. M.; Gabbay, K. H.; Quiocho, F. A., An unlikely sugar substrate site in the 1.65 A structure of the human aldose reductase holoenzyme implicated in diabetic complications. *Science* **1992,** *257* (5066), 81-4.

292. Feldman, H. B.; Szczepanik, P. A.; Havre, P.; Corrall, R. J.; Yu, L. C.; Rodman, H. M.; Rosner, B. A.; Klein, P. D.; Landau, B. R., Stereospecificity of the hydrogen transfer catalyzed by human placental aldose reductase. *Biochimica et biophysica acta* **1977,** *480* (1), 14-20.

293. Liu, S. Q.; Bhatnagar, A.; Srivastava, S. K., Bovine lens aldose reductase. pH-dependence of steady-state kinetic parameters and nucleotide binding. *The Journal of biological chemistry* **1993,** *268* (34), 25494-9.

294. Bohren, K. M.; Grimshaw, C. E.; Lai, C. J.; Harrison, D. H.; Ringe, D.; Petsko, G. A.; Gabbay, K. H., Tyrosine-48 is the proton donor and histidine-110 directs substrate stereochemical selectivity in the reduction reaction of human aldose reductase: enzyme kinetics and crystal structure of the Y48H mutant enzyme. *Biochemistry* **1994,** *33* (8), 2021-32.

295. Lee, Y. S.; Hodoscek, M.; Brooks, B. R.; Kador, P. F., Catalytic mechanism of aldose reductase studied by the combined potentials of quantum mechanics and molecular mechanics. *Biophysical chemistry* **1998,** *70* (3), 203-16.

296. Varnai, P.; Richards, W. G.; Lyne, P. D., Modelling the catalytic reaction in human aldose reductase. *Proteins* **1999,** *37* (2), 218-27.

297. Várnai, P.; Warshel, A., Computer Simulation Studies of the Catalytic Mechanism of Human Aldose Reductase. *Journal of the American Chemical Society* **2000,** *122* (16), 3849-3860.

298. Grimshaw, C. E.; Bohren, K. M.; Lai, C. J.; Gabbay, K. H., Human aldose reductase: rate constants for a mechanism including interconversion of ternary complexes by recombinant wild-type enzyme. *Biochemistry* **1995,** *34* (44), 14356-65.

299. Lodola, A.; De Vivo, M., The increasing role of QM/MM in drug discovery. *Advances in protein chemistry and structural biology* **2012,** *87*, 337-62.

300. Ullmann, G. M.; Knapp, E. W., Electrostatic models for computing protonation and redox equilibria in proteins. *European biophysics journal : EBJ* **1999,** *28* (7), 533-51.

301. Uranga, J.; Mikulskis, P.; Genheden, S.; Ryde, U., Can the protonation state of histidine residues be determined from molecular dynamics simulations? *Computational and Theoretical Chemistry* **2012,** *1000*, 75-84.

302. Alexov, E.; Mehler, E. L.; Baker, N.; Baptista, A. M.; Huang, Y.; Milletti, F.; Nielsen, J. E.; Farrell, D.; Carstensen, T.; Olsson, M. H.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, J. M., Progress in the prediction of pKa values in proteins. *Proteins* **2011,** *79* (12), 3260-75.

303. Olsson, M. H.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H., PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput* **2011,** *7* (2), 525-37.

304. MacroModel, v., Schrödinger, LLC, New York, NY, 2015.

305. Bruno, I. J.; Cole, J. C.; Kessler, M.; Luo, J.; Motherwell, W. D.; Purkis, L. H.; Smith, B. R.; Taylor, R.; Cooper, R. I.; Harris, S. E.; Orpen, A. G., Retrieval of crystallographically-derived molecular geometry information. *Journal of chemical information and computer sciences* **2004,** *44* (6), 2133-44.

306. Impact, v., Schrödinger, LLC, New York, NY, 2014.

307. Polak, E.; Ribiere, G., Note sur la convergence de méthodes de directions conjuguées. *R.I.R.O.* **1969,** *3* (16), 35-43.

308. El-Kabbani, O.; Ruiz, F.; Darmanin, C.; Chung, R. P., Aldose reductase structures: implications for mechanism and inhibition. *Cellular and molecular life sciences : CMLS* **2004,** *61* (7-8), 750-62.

309. Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, T. J., How Large Should the QM Region Be in QM/MM Calculations? The Case of Catechol O-Methyltransferase. *The Journal of Physical Chemistry B* **2016,** *120* (44), 11381-11394.

310. Liao, R. Z.; Thiel, W., Convergence in the QM-only and QM/MM modeling of enzymatic reactions: A case study for acetylene hydratase. *J Comput Chem* **2013,** *34* (27), 2389-97.

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

312. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P., Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society* **1985,** *107* (13), 3902-3909.

313. Del Corso, A.; Cappiello, M.; Mura, U., From a dull enzyme to something else: facts and perspectives regarding aldose reductase. *Curr Med Chem* **2008,** *15* (15), 1452-61.

314. Tarle, I.; Borhani, D. W.; Wilson, D. K.; Quiocho, F. A.; Petrash, J. M., Probing the active site of human aldose reductase. Site-directed mutagenesis of Asp-43, Tyr-48, Lys-77, and His-110. *The Journal of biological chemistry* **1993,** *268* (34), 25687-93.

315. Ruiz, F.; Hazemann, I.; Mitschler, A.; Joachimiak, A.; Schneider, T.; Karplus, M.; Podjarny, A., The crystallographic structure of the aldose reductase-IDD552 complex shows direct proton donation from tyrosine 48. *Acta crystallographica. Section D, Biological crystallography* **2004,** *60* (Pt 8), 1347-54.

316. Blakeley, M. P.; Ruiz, F.; Cachau, R.; Hazemann, I.; Meilleur, F.; Mitschler, A.; Ginell, S.; Afonine, P.; Ventura, O. N.; Cousido-Siah, A.; Haertlein, M.; Joachimiak, A.; Myles, D.; Podjarny, A., Quantum model of catalysis based on a mobile proton revealed by subatomic x-ray and neutron diffraction studies of h-aldose reductase. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** *105* (6), 1844-8.

317. Pattabiraman, V. R.; Bode, J. W., Rethinking amide bond synthesis. *Nature* **2011,** *480* (7378), 471-9.

318. Carey, J. S.; Laffan, D.; Thomson, C.; Williams, M. T., Analysis of the reactions used for the preparation of drug candidate molecules. *Organic & biomolecular chemistry* **2006,** *4* (12), 2337-47.

319. Roughley, S. D.; Jordan, A. M., The medicinal chemist's toolbox: an analysis of reactions used in the pursuit of drug candidates. *Journal of medicinal chemistry* **2011,** *54* (10), 3451-79.

320. Charville, H.; Jackson, D. A.; Hodges, G.; Whiting, A.; Wilson, M. R., The Uncatalyzed Direct Amide Formation Reaction – Mechanism Studies and the Key Role of Carboxylic Acid H-Bonding. *European Journal of Organic Chemistry* **2011,** *2011* (30), 5981-5990.

321. Valeur, E.; Bradley, M., Amide bond formation: beyond the myth of coupling reagents. *Chemical Society reviews* **2009,** *38* (2), 606-31.

322. Ali, Y. B.; Verger, R.; Abousalham, A., Lipases or esterases: does it really matter? Toward a new bio-physico-chemical classification. *Methods in molecular biology (Clifton, N.J.)* **2012,** *861*, 31-51.

323. Hofmann, K.; Bucher, P.; Falquet, L.; Bairoch, A., The PROSITE database, its status in 1999. *Nucleic acids research* **1999,** *27* (1), 215-9.

324. Casas-Godoy, L.; Duquesne, S.; Bordes, F.; Sandoval, G.; Marty, A., Lipases: An Overview. In *Lipases and Phospholipases: Methods and Protocols*, Sandoval, G., Ed. Humana Press: Totowa, NJ, 2012; pp 3-30.

325. Goswami, A.; Van Lanen, S. G., Enzymatic strategies and biocatalysts for amide bond formation: tricks of the trade outside of the ribosome. *Molecular bioSystems* **2015,** *11* (2), 338-53.

326. Romano, D.; Bonomi, F.; de Mattos, M. C.; de Sousa Fonseca, T.; de Oliveira Mda, C.; Molinari, F., Esterases as stereoselective biocatalysts. *Biotechnology advances* **2015,** *33* (5), 547-65.

327. Pitzer, J.; Steiner, K., Amides in Nature and Biocatalysis. *Journal of biotechnology* **2016,** *235*, 32-46.

328. Down, K.; Amour, A.; Baldwin, I. R.; Cooper, A. W. J.; Deakin, A. M.; Felton, L. M.; Guntrip, S. B.; Hardy, C.; Harrison, Z. A.; Jones, K. L.; Jones, P.; Keeling, S. E.; Le, J.; Livia, S.; Lucas, F.; Lunniss, C. J.; Parr, N. J.; Robinson, E.; Rowland, P.; Smith, S.; Thomas, D. A.; Vitulli, G.; Washio, Y.; Hamblin, J. N., Optimization of Novel Indazoles as Highly Potent and Selective Inhibitors of Phosphoinositide 3-Kinase δ for the Treatment of Respiratory Disease. *Journal of medicinal chemistry* **2015,** *58* (18), 7381-7399.

329. Patent Application WO/2018/029126.

330. De Simone, G.; Galdiero, S.; Manco, G.; Lang, D.; Rossi, M.; Pedone, C., A snapshot of a transition state analogue of a novel thermophilic esterase belonging to the subfamily of mammalian hormone-sensitive lipase. *Journal of molecular biology* **2000,** *303* (5), 761-71.

331. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory and Computation* **2016,** *12* (1), 281-296.

332. Cooper, A. M.; Kastner, J., Averaging techniques for reaction barriers in QM/MM simulations. *Chemphyschem : a European journal of chemical physics and physical chemistry* **2014,** *15* (15), 3264-9.

333. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *Journal of molecular biology* **1990,** *215* (3), 403-10.

334. Consortium, U., UniProt: a hub for protein information. *Nucleic acids research* **2015,** *43* (Database issue), D204-12.

335. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; Thompson, J. D.; Higgins, D. G., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **2011,** *7*, 539.

336. Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E., WebLogo: a sequence logo generator. *Genome research* **2004,** *14* (6), 1188-90.

337. Arnold, F. H., Directed Evolution: Bringing New Chemistry to Life. *Angewandte Chemie (International ed. in English)* **2017**.

338. Zhang, L.; Gao, B.; Yuan, Z.; He, X.; Yuan, Y. A.; Zhang, J. Z. H.; Wei, D., Structure, mechanism, and enantioselectivity shifting of lipase LipK107 with a simple way. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2014,** *1844* (7), 1183-1192.

339. Korman, T. P.; Bowie, J. U., Crystal Structure of Proteus mirabilis Lipase, a Novel Lipase from the Proteus/Psychrophilic Subfamily of Lipase Family I.1. *PLOS ONE* **2012,** *7* (12), e52890.

340. Korman, T. P.; Sahachartsiri, B.; Charbonneau, D. M.; Huang, G. L.; Beauregard, M.; Bowie, J. U., Dieselzymes: development of a stable and methanol tolerant lipase for biodiesel production by directed evolution. *Biotechnology for Biofuels* **2013,** *6* (1), 70.

341. Nardini, M.; Lang, D. A.; Liebeton, K.; Jaeger, K. E.; Dijkstra, B. W., Crystal structure of pseudomonas aeruginosa lipase in the open conformation. The prototype for family I.1 of bacterial lipases. *The Journal of biological chemistry* **2000,** *275* (40), 31219-25.

342. Reis, P.; Holmberg, K.; Watzke, H.; Leser, M. E.; Miller, R., Lipases at interfaces: A review. *Advances in Colloid and Interface Science* **2009,** *147-148*, 237-250.

343. Schrödinger Release 2016-1: LigPrep, S., LLC, New York, NY, 2016.

344. Jaguar, v., Schrödinger, LLC, New York, NY, 2016.

345. Jorgensen, W. L.; Laird, E. R.; Nguyen, T. B.; Tirado-Rives, J., Monte Carlo simulations of pure liquid substituted benzenes with OPLS potential functions. *J. Comput. Chem.* **1993,** *14* (2), 206-215.

346. Edgar, R. C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **2004,** *32* (5), 1792-1797.

347. Bornscheuer, U. T., The fourth wave of biocatalysis is approaching. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2018,** *376* (2110).

348. Crow, J. M., Go with the fold. *Chemistry World* 30 March, 2018.

349. Moult, J.; Fidelis, K.; Kryshtafovych, A.; Schwede, T.; Tramontano, A., Critical assessment of methods of protein structure prediction (CASP)—Round XII. *Proteins: Structure, Function, and Bioinformatics* **2017,** *86* (S1), 7-15.

350. Svensson, F.; Engen, K.; Lundbäck, T.; Larhed, M.; Sköld, C., Virtual Screening for Transition State Analogue Inhibitors of IRAP Based on Quantum Mechanically Derived Reaction Coordinates. *Journal of Chemical Information and Modeling* **2015,** *55* (9), 1984-1993.

351. Schramm, V. L., Enzymatic transition states and transition state analogues. *Current opinion in structural biology* **2005,** *15* (6), 604-613.

352. Kästner, J., Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011,** *1* (6), 932-942.

353. Isborn, C. M.; Götz, A. W.; Clark, M. A.; Walker, R. C.; Martínez, T. J., Electronic Absorption Spectra from MM and ab Initio QM/MM Molecular Dynamics: Environmental Effects on the Absorption Spectrum of Photoactive Yellow Protein. *Journal of Chemical Theory and Computation* **2012,** *8* (12), 5092-5106.

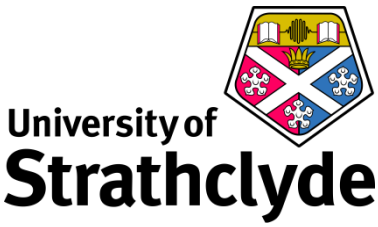
354. Mazanetz, M. P.; Marmon, R. J.; Reisser, C. B.; Morao, I., Drug discovery applications for KNIME: an open source data mining platform. *Current topics in medicinal chemistry* **2012,** *12* (18), 1965-79.

355. Sirin, S.; Pearlman, D. A.; Sherman, W., Physics‐based enzyme design: Predicting binding affinity and catalytic activity. *Proteins: Structure, Function, and Bioinformatics* **2014,** *82* (12), 3397-3409.

356. Jacobson, L. D.; Bochevarov, A. D.; Watson, M. A.; Hughes, T. F.; Rinaldo, D.; Ehrlich, S.; Steinbrecher, T. B.; Vaitheeswaran, S.; Philipp, D. M.; Halls, M. D.; Friesner, R. A., Automated Transition State Search and Its Application to Diverse Types of Organic Reactions. *Journal of Chemical Theory and Computation* **2017,** *13* (11), 5780-5797.

357. Kumar, A.; Dhar, K.; Kanwar, S. S.; Arora, P. K., Lipase catalysis in organic solvents: advantages and applications. *Biological Procedures Online* **2016,** *18* (1), 2.

358. Rehm, S.; Trodler, P.; Pleiss, J., Solvent-induced lid opening in lipases: a molecular dynamics study. *Protein science : a publication of the Protein Society* **2010,** *19* (11), 2122-30.



Molecular Design, GlaxoSmithKline

Department of Pure and Applied Chemistry

APPENDICES: COMPUTATIONAL MODELLING OF ENZYME ACTIVITY TO SPEED UP BIOCATALYST REDESIGN

by

MARIE-PIERRE DRÉANIC

A thesis submitted to the Department of Pure and Applied Chemistry, University of Strathclyde, in part fulfillment of the regulations for the degree of Doctor of Philosophy in Chemistry.

April 2018

Table of Contents

1. Appendix for Chapter 4

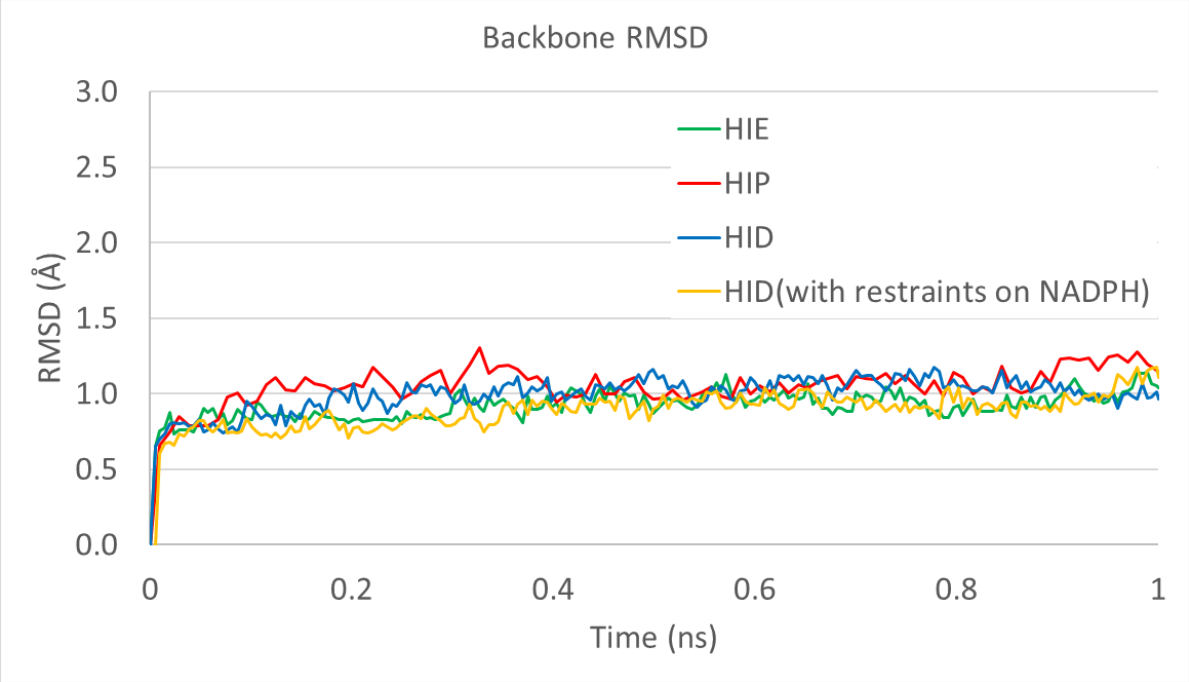


Figure A1.1. Coordinates RMSD of MD-simulated for aldose reductase in water and for the three protonation models: HIP (red), HIE (green) and HID (blue). An additional HID simulation was doned with restrains on NADPH (yellow). The RMSD is calculated on the protein backbone.

Protonated HIP110: other snapshots

Snapshot 2

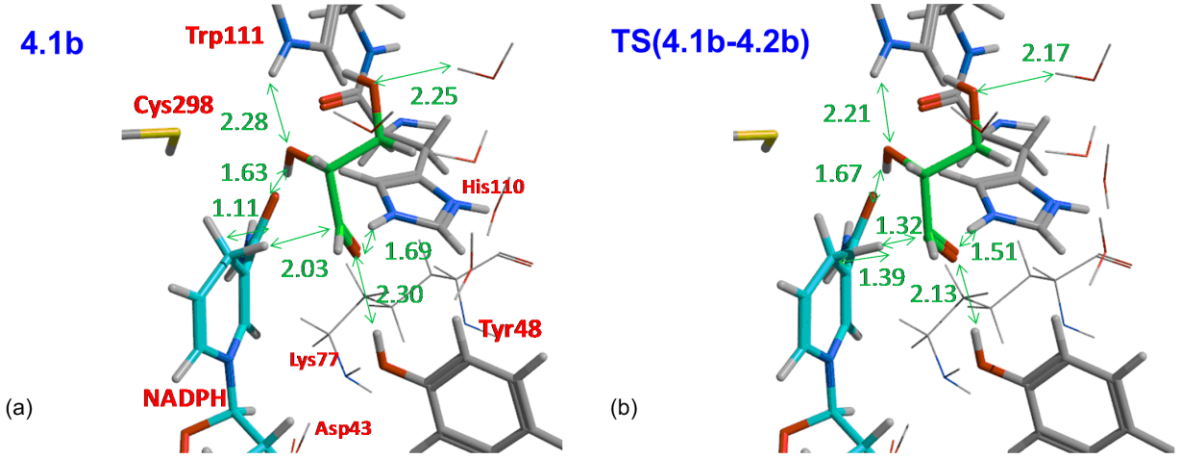


Figure A1.2. Reaction intermediates of the GLD reduction by AR with HIP110 as studied by the QM/MM model for snapshot 2 of HIP (a) Enzyme-substrate complex (b) Transition state (Distances shown in green, atoms numbers in brown and residues names in red).

Unprotonated HIE110: other snapshots

Snapshot 2

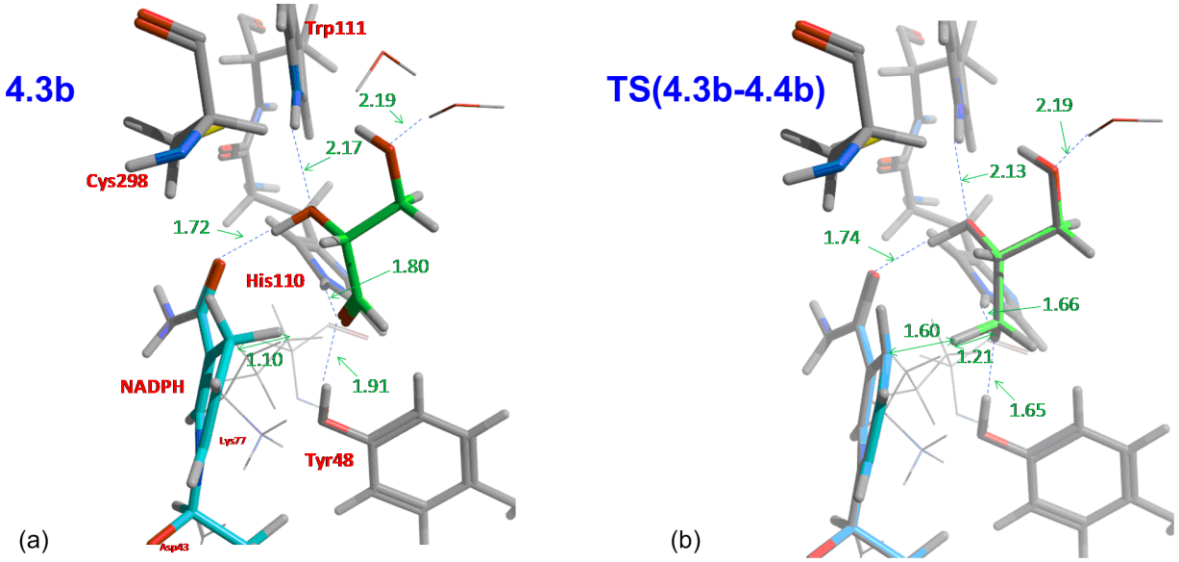


Figure A1.3. Reaction intermediates of the GLD reduction by AR with HIE110 as studied by the QM/MM model for snapshot 2 of HIE. (a) Enzyme-substrate complex (b) Transition state 1 (Distances shown in green, atoms numbers in brown and residues names in red).

Snapshot 3

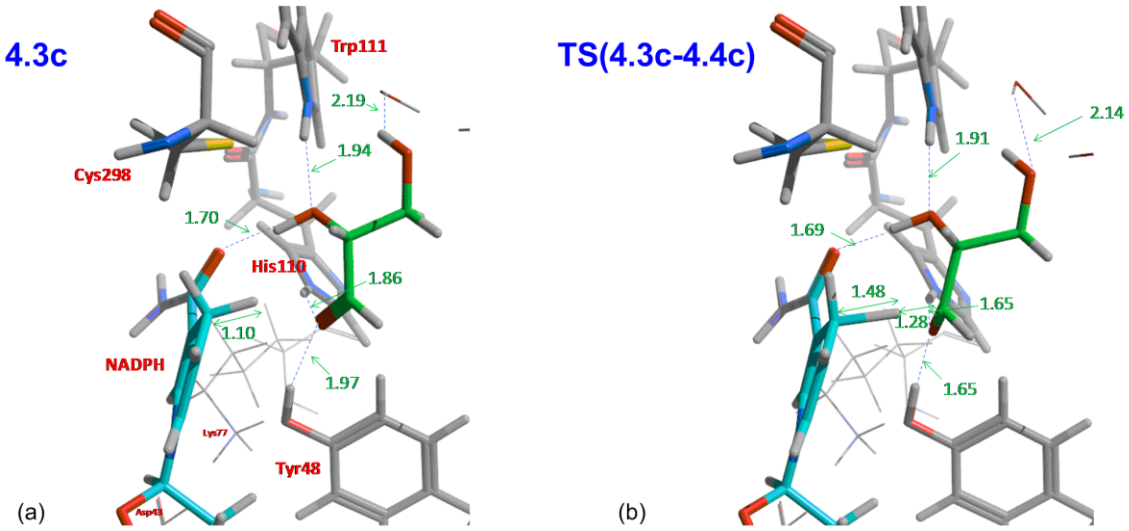
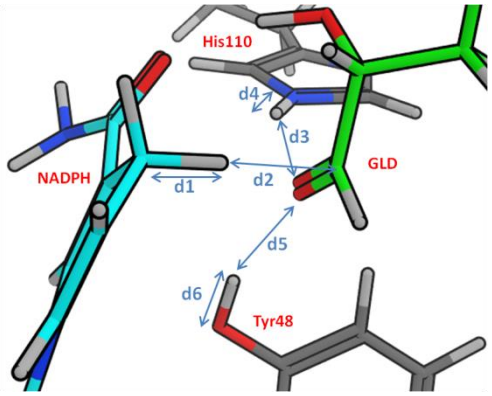


Figure A1.4. Reaction intermediates of the GLD reduction by AR with HIE110 as studied by the QM/MM model for snapshot 3 of HIE. (a) Enzyme-substrate complex (b) Transition state 1 (Distances shown in green, atoms numbers in brown and residues names in red).

Effect of the basis set and the QM size on selected geometrical characteristics

Although the energies differ, the geometrical characteristics were very similar to previous studies. For the HIP model, the geometries look particularly similar at the transition state (**Table A1.1a**). At the reactant state the ligand seem to be in closer proximity to the enzyme in our model as both hydrogen bonds between the oxygen carbonyl and residues His110 and Tyr48 are shorter [1.83 Å and 2.02 Å, respectively], compared to the Lee and co-workers[1] model [2.0 Å and 2.53 Å, respectively]. The hydride is also initially closer to the carbonyl carbon in our model compared to previous results [2 Å and 2.59 Å, respectively]. The same is also observed for the HIE model (**Table A1.1b**). At the reactant state the ligand seem to be closer proximity to the enzyme in our model as both hydrogen bonds between the oxygen carbonyl and residues His110 and Tyr48 are shorter [1.86 Å and 1.89 Å, respectively], compared to the Varnai and co-workers[2] model [2.22 Å and 2.64 Å, respectively]. The hydride is also initially closer to the carbonyl carbon in our model compared to previous results [2.23 Å and 2.49 Å, respectively].

Table A1.1. Comparison of key distances from Lee and co-workers[1] model (Lee) and Varnai and co-workers[2] model (Var.) to (a) HIP and (b) HIE models. Δ is the difference between the two results. A threshold of 0.1 Å has established in order to differentiate between small changes (green) and bigger ones(red).



a.

	reactant			TS			product		
	HIP	Lee	Δ	HIP	Lee	Δ	HIP	Lee	Δ
distances between atoms in the active site (Å)									
d1	1.11	1.09	-0.02	1.45	1.35	-0.1	3.22	NC	/
d2	2.00	2.59	0.59	1.26	1.34	0.08	1.09	NC	/
d3	1.83	2.00	0.17	1.61	1.57	-0.04	1.01	NC	/
d4	1.03	1.00	-0.03	1.07	1.06	-0.01	1.72	1.96	0.24
d5	2.02	2.53	0.51	1.83	1.75	-0.08	1.72	1.91	0.19
d6	0.98	0.96	-0.02	1.00	0.98	-0.02	1.00	0.97	-0.03

b.

	R			TS1			I			TS2			P		
	HIE	Var.	Δ	HIE	Var.	Δ	HIE	Var.	Δ	HIE	Var.	Δ	HIE	Var.	Δ
distances between qm atoms in the active site (Å)															
d1	1.10	1.13	0.03	1.54	1.75	0.21	1.96	2.37	0.41	2.22	2.38	0.16	2.49	3.42	0.93
d2	2.23	2.49	0.26	1.34	1.2	-0.14	1.13	1.14	0.01	1.11	1.14	0.03	1.10	1.12	0.02
d3	1.86	2.22	0.36	1.69	1.99	0.3	1.62	1.9	0.28	1.62	1.93	0.31	1.70	2.12	0.42
d4	1.02	1.00	-0.02	1.04	1.01	-0.03	1.06	1.02	-0.04	1.05	1.01	-0.04	1.03	1.00	-0.03
d5	1.89	2.64	0.75	1.66	1.86	-0.2	1.48	1.8	0.32	1.36	1.48	0.12	1.01	0.97	-0.04
d6	0.98	0.97	-0.01	1.02	0.98	-0.04	1.06	0.99	-0.07	1.13	1.06	-0.07	1.71	2.11	0.4

2. Appendix for Chapter 5

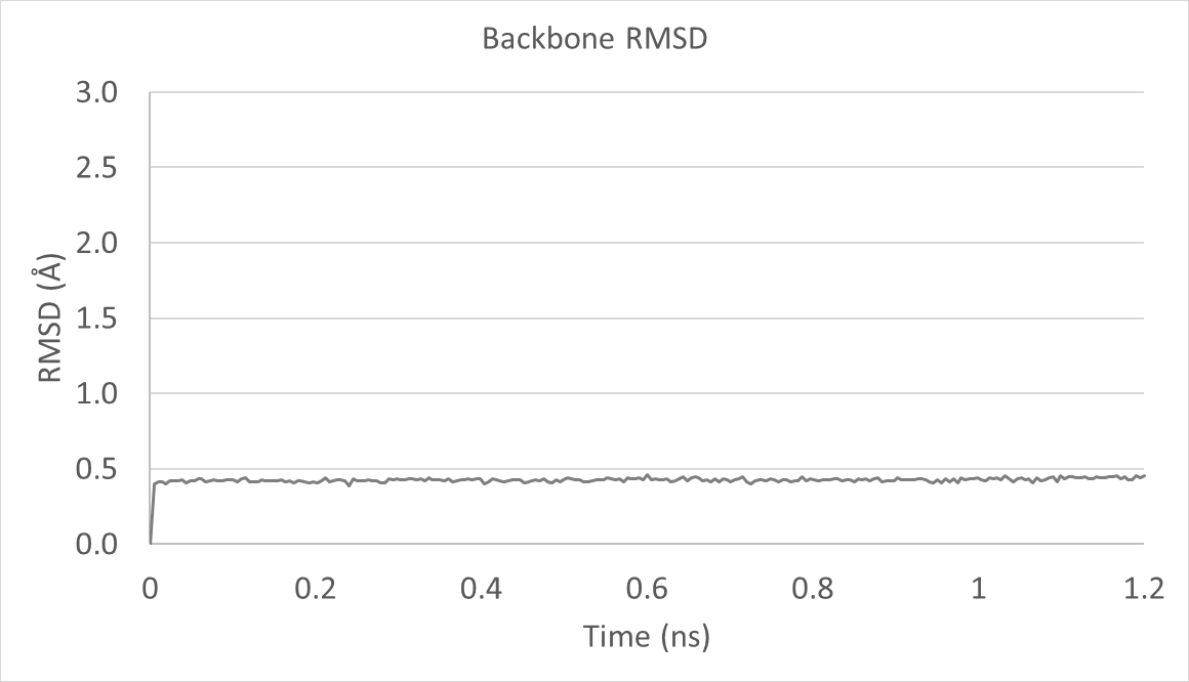


Figure A2.1. Coordinates RMSD of MD-simulated for the carboxylesterase in water. The RMSD is calculated on the protein backbone.

3. Appendix for Chapter 6

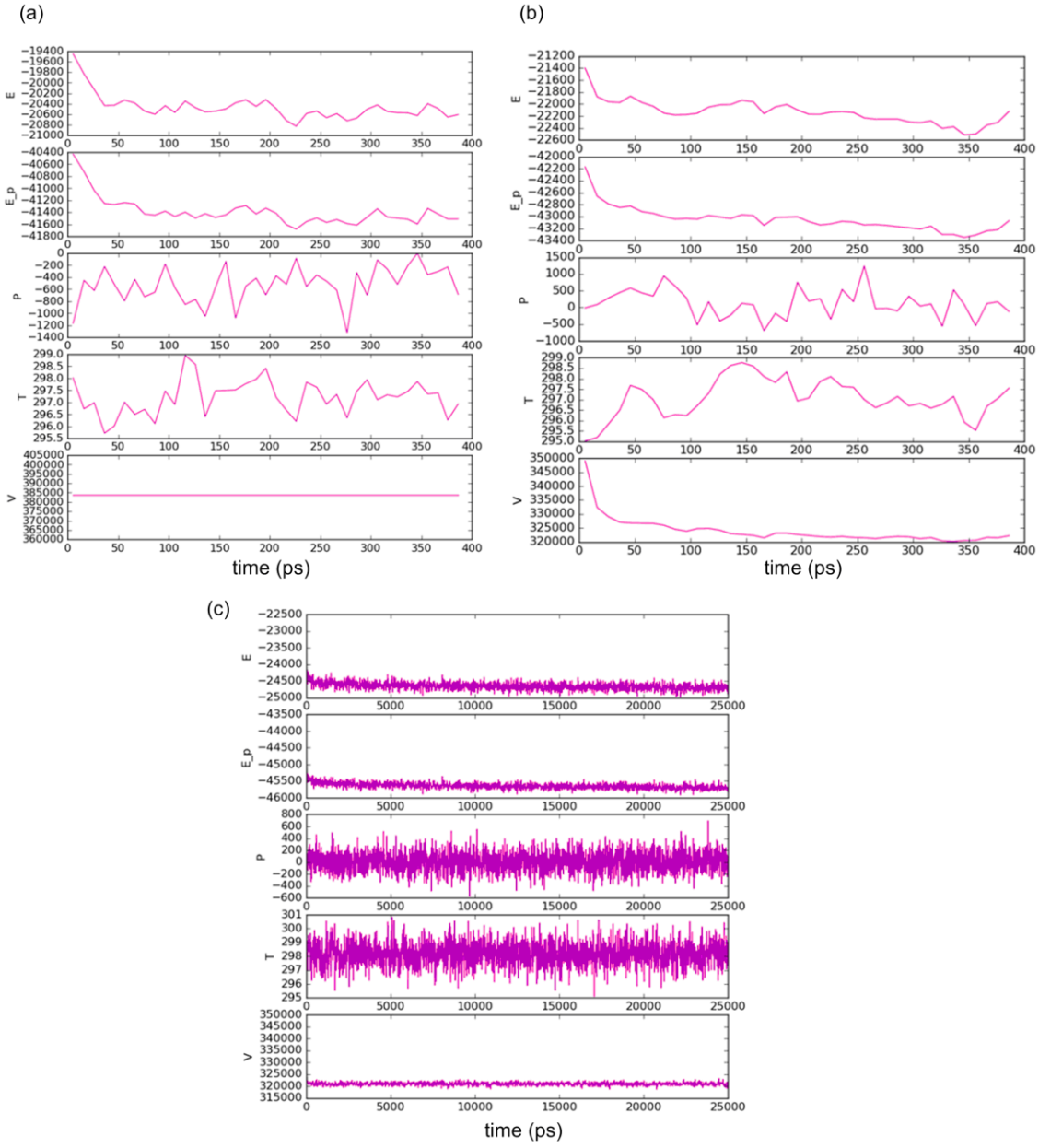


Figure A3.1. Evolution of the total energy (E, in kcal/mol), potential energy (E_p, in kcal/mol), pressure (P, in Bar), temperature (T, in Kelvin) and volume (V, in $Å^3$) over time during the equilibration protocol for the toluene model described in Figure 6.3 of the

thesis. (a) 400 ps NVT simulation, (b) 400ps NPT simulation, (c) 25 ns NPT production run.

References

1. Lee, Y. S.; Hodoscek, M.; Brooks, B. R.; Kador, P. F., Catalytic mechanism of aldose reductase studied by the combined potentials of quantum mechanics and molecular mechanics. *Biophysical chemistry* **1998,** *70* (3), 203-16.

2. Varnai, P.; Richards, W. G.; Lyne, P. D., Modelling the catalytic reaction in human aldose reductase. *Proteins* **1999,** *37* (2), 218-27.