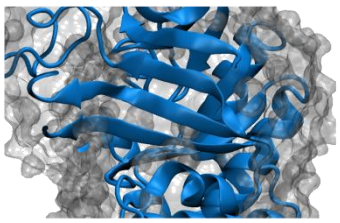
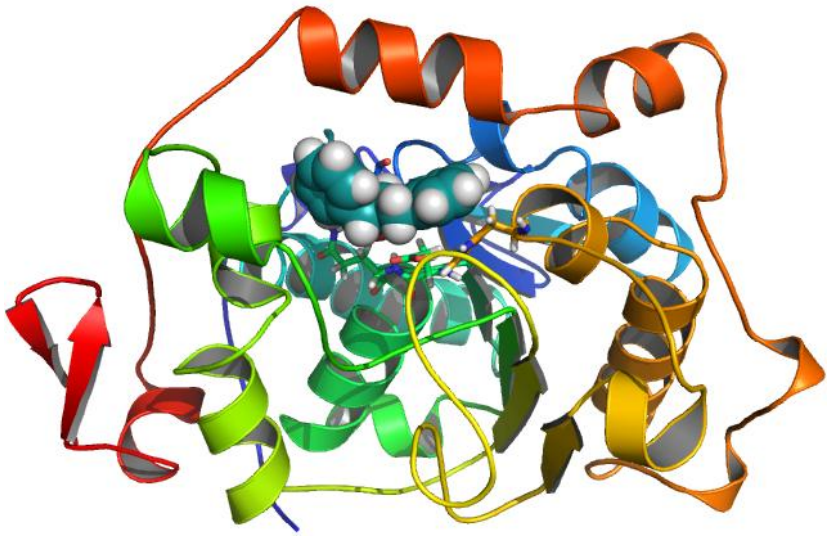
BioComp



UNIVERSITAT JAUME·I

Computational Design of New Epoxidase Based on Promiscuous CAL-B Scaffold



José Recatalà Gómez

Supervisor: Prof. Dr. Vicente Moliner Ibáñez

CONTENTS

1. - Introduction

1.1. - Biochemistry fundamentals

1.1.1. - Enzymatic catalysis

There are two fundamental conditions for life. First, the organism must be able to self-replicate; second, it must be able to catalyse chemical reactions efficiently and selectively. Without catalysis, the very basic and important chemical reactions for life, such as conversion of sugars in CO_2 and H_2O, could not occur on a useful time scale, and thus could not sustain life[1].

Enzymes are generally proteins that lead, control and speed up the chemical reactions in biological systems[2].In fact, this control can be carried out not only catalyzing the main reaction, but inhibiting the side reactions that can take place between the reactants or between the reactants and the solvent[2].

Enzymes catalyze eco-friendly organic processes, generally under mild reaction conditions (pH, pressure, temperature...) showing high stability and catalytic efficiency[3], broad substrate specifity and tolerance to both aqueous and organic media for a broad range of biocatalysts[4].

Thanks to all the previously mentioned factors, enzymes have perfectly fit within the basic principles of sustainable chemistry[5], being able to perform regio-, chemo- and stereoselective processes[6].

Because all of the previously mentioned facts, biocatalysis has emerged as an elegant synthetic methodology that was at the beginning not so accepted, as there was a bit of concern from the members of the scientific community. Despite these initial concerns, biocatalysis is now very popular and widely used in organic synthesis.[7]

An enzyme is usually a protein (though it can also be a nucleic acid) built up of 20 different species of amino acids linked together to form one or more long chains, chained between peptidic bonds. Its molecular weights range from ten thousand to hundreds of thousands of Dalton[8]. The long chains of which they are constituted must be folded in such way that a three-dimensional pocket or cleft is created. It is in this cleft where the interacting compounds (known as substrates) fit in a very precise way[8]. This part of the enzyme where the substrate binds to and where the catalytic process takes place is known as active site[9, 10].

Some enzymes require cofactors to be active. The cofactors are non-proteic components which can be sorted into:

- Metal Ions: Some common metal-ionic cofactors are Zn^{2+}, Cu^{2+}, Mg^{2+} amongst others.
- Organic cofactors: Also known as coenzymes[10].

The latter, at the same time, can be co-substrates, which are only transiently associated with the enzyme (NAD^+, $NADP^+$...) or prosthetic groups, which are permanently bounded to the enzyme, often by covalent bonds (heme group).

Summarizing, the inactive apoenzyme bounds with the cofactor to form a holoenzyme, which is already active.

It must be borne in mind that, in order to complete a catalytic cycle, the enzyme must return to its original state.

The most remarkable properties of enzymes as catalysts are[11, 12]:

-*Catalytic power:* The rate of a non-enzymatic reaction[13] varies from 10^{-7} to 10^{-1} s^{-1}, while in an enzymatic reaction it can be in a range from 10 to 10^{7} s^{-1}.

-*Specifity:* Enzymes have usually a vast degree of specificity with respect to their substrates and products. Consequently, the amount of side products is nearly zero.

-*Milder reaction conditions*: Enzymes usually work under neutral pH, mild temperature and atmospheric pressure.

-*Capacity for regulation:* The activity of some enzymes can vary in response to the concentration of substances other than their substrate.

Each enzyme has two names and classification numbers[14]: Trivial or old name; official systematic name which is the name of its substrate or substrates followed by a word ending in *–ase*, specifying the type of reaction that the enzyme catalyzes, and classification number from the Enzyme Commission. In the present work, the enzyme which is studied is the *Candida Antarctica Lipase B (CAL-B)*, which code is EC3.1.1.3:

3: Hydrolases

3.1: Acting on ester bonds

3.1.1: Carboxylic ester hydrolases

3.1.1.3: Triacylglycerol lipase

It is also interesting to remark that both pH and temperature have a very important effect on enzymatic catalysis. Most of the enzymes have a characteristic pH (usually neutral pH) where they present the largest activity. Catalytic activity and pH are related with the acid-base behaviour of the enzyme and the substrate amongst many other factors.

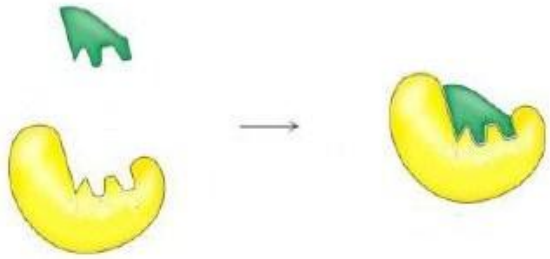
On the other hand, for exothermic chemical reactions, the higher the temperature, the larger the rate reaction value of the chemical reaction. However, due to the fact that enzymes suffer of denaturalization phenomena at higher temperatures, they also have a certain value of temperature, also known as 'optimal temperature' where they show the highest activity value.

1.1.2. - Models and mathematical descriptions in enzymology

The word enzyme used for the first time in 1878 by Kühne and its meaning is 'in yeast', as a try to emphasize that the catalytic activity was the manifestation of an extract or secretion, rather than a manifestation of the entire organism.

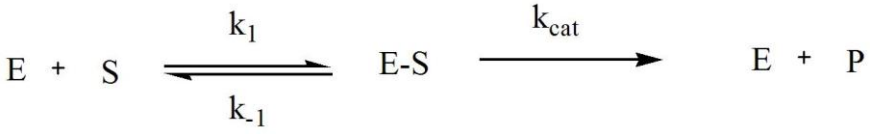
In 1894, Fischer proposed the *lock & key* model[15]. The mainstream of this mechanism is that both the structure of the enzyme and the structure of substrate are complementary and completely rigid. The following figure[16] depicts the *lock & key* model:

Figure 1: Schematic representation of the lock & key model



Later on, in 1913 a new model for enzymatic catalysis was suggested by Leonor Michaelis and Maud Menten[17], based on the following catalytic scheme (for single-substrate enzyme-catalyzed reaction):

Scheme 1: Schematic representation of the Michaelis-Menten catalytic reaction



According to the Scheme 1, the enzyme-substrate complex, E-S, also known as the Michaelis-Menten's coordination complex is formed in the first reversible step. Then comes a slow step where the complex goes through different chemical transformations, giving the products od the reaction, P, and releasing the enzyme, E.

The reaction rate is usually written as:

$$r_{cat} = \frac{r_{max} \cdot [S]}{K_M + [S]} \qquad (1)$$

The equation (1) is also known as the Michaelis-Menten equation, where

r_{max}: is the maximum speed.

K_M: concentration of substrate at which the reaction rate is half of the maximum rate.

[S]: concentration of substrate.

In 1930, Haldane[18] proposed the induced-fit hypothesis: the active site of the enzyme adopts an optimal conformation to interact with the substrate, but only when the substrate itself is present. The following figure [16] depicts the induced-fit model:

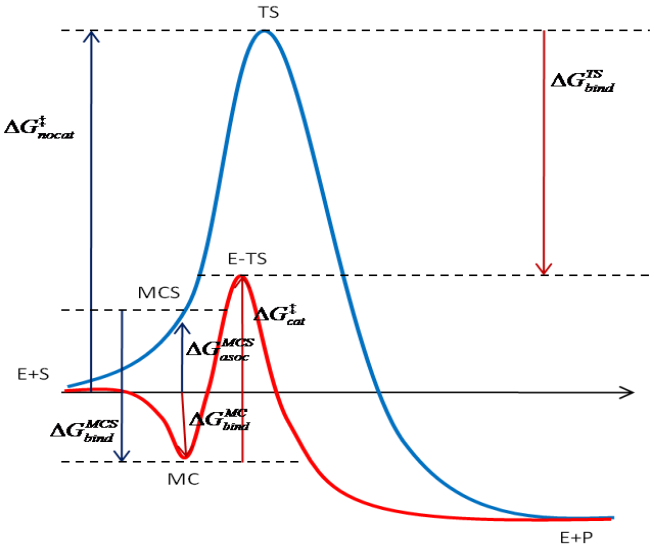
Figure 2: Schematic representation of the induced-fit model

In 1946, the preferential binding to the transition state (TS) in a catalytic reaction was proposed by Pauling[19]. An enzyme will bind preferentially, and with greater affinity to the TS of a catalytic reaction rather than to its substrate or products. Hence, the active site of the enzyme is not complementary to the substrate in its fundamental state but to the TS.

What makes the enzymes such interesting molecules, is the fact that the enzymes modify the rates of the reactions but not the equilibriums of them. This is translated into a higher rate with a lowering of the energy of activation (E_a).

The energetic profile of the reaction for both a catalyzed and a non-catalyzed reaction is depicted in the following figure, by the correspondence of the former to the reaction described in the Scheme 1:

Figure 3: Free energy profile for both an enzyme catalyzed reaction (in red) and a non-catalyzed reaction (in blue). Figure from reference[20].



The non-catalyzed reaction takes place in just one step, with a value of activation energy of $\Delta G_{uncat}^{\ddagger}$ whilst the catalyzed reaction, in this example, proceeds in two steps:

- First step: the Michaelis-Menten complex (MC) i.e. the substrate-enzyme complex is formed by the loss of binding energy ΔG_{bind}^{MC}.
- Second step: The chemical reaction takes place, with an energetic cost $\Delta G_{cat}^{\ddagger}$.

As it can be seen by simple inspection of the Figure 3, $\Delta G_{uncat}^{\ddagger} > \Delta G_{cat}^{\ddagger}$.

The, the reaction rate constant, k_{cat}, for the catalytic reaction can be obtained by using the Transition State Theory:

$$k_{cat} = \frac{k_B T}{h} \exp\left\{\frac{-\Delta G_{cat}^{\ddagger}}{RT}\right\} \qquad (2)$$

Where:

k_B is the Boltzmann constant, which value in I.S. units is $1{,}3806488(13)\cdot 10^{-23}$ $J\cdot K^{-1}$.

T is the temperature of the system

h is the Planck constant, value in I.S. units is $6.62606957(29) \times 10^{-34}$ J·s.

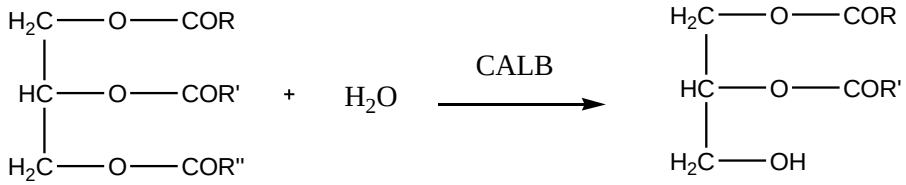
$\Delta G_{cat}^{\ddagger}$ is the free energy of the catalytic reaction.

R is the gas constant, which value in the I.S. units is 8.3144621(75) $J\ K^{-1}\ mol^{-1}$.

1.1.3.- Hydrolases. *Candida Antarctica Lipase B* (CAL-B)

Candida Antarctica Lipase B, also known as CAL-B, is a monomeric serine hydrolase whose mechanism of primary reaction, which is the hydrolysis of ester bonds, was recently described by using both experimental[21] and theoretical[22] methods, providing interesting insights into the functions of residues in the active site. The following scheme depicts the primary activity of CAL-B:

Scheme 2: Schematic representation of primary reaction of CAL-B



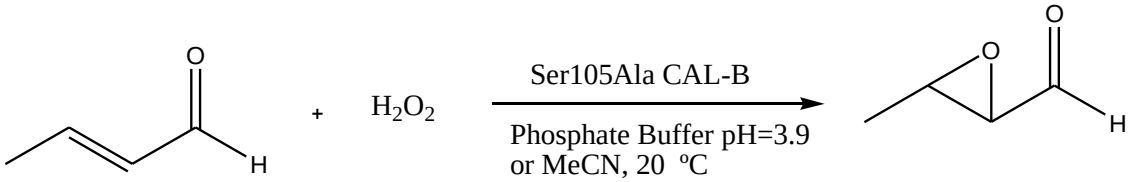
In addition, it also shows catalytic activity for a plethora of reactions, like carboxylic acid esterase, thioesterase, peptidase, dehalogenase, epoxide hydrolase, halo peroxidase, and cleaving C-C bonds[23].

Henceforth, CAL-B has been widely studied because of its high catalytic promiscuity[24]. This promiscuous activity shown by CAL-B has proven to be a versatile catalyst for a wide range of biotransformations[25]. CAL-B has also been used in material science, for synthesis of hierarchical materials, as catalyst in ring opening polymerization reactions when attached to porous materials like another polymers or nanostructures[26].

Introductory courses in biochemistry teach that enzymes are specific for their substrates and the reactions they catalyze. Enzymes diverging from this statement are sometimes called

promiscuous[27]. Enzyme catalytic promiscuity refers to the ability of an enzyme active site to catalyze more than one different chemical transformation.
An example of this promiscuous behaviour in CAL-B may be the one found by Berglung and his research group. They discovered that both wild-type CAL-B and Ser105Ala variant were able to catalyze the direct epoxidation of α/β-unsaturated compounds with hydrogen peroxide in both aqueous and organic media[28].

Scheme 3: Schematic representation of CAL-B catalyzed direct epoxidation of a α/β-unsaturated aldehyde



In this thesis, we focus on epoxide hydrolases[29] (EHs, E.C. 3.3.2.3) whose main activity is the transformation an epoxide into the corresponding 1,2-diol using a water molecule as the only cosubstrate. The role of epoxide hydrolases seems to differ profoundly from organism to organism. Overall, these enzymes have three main functions: detoxification, catabolism, and regulation of signalling molecules. However, EHs have been mainly source of interest because of their potential applications in chiral chemistry[30].

There are several EHs but one, the most studied is the soluble Epoxide Hydrolase (sEH).

sEH hydrolyses a broad range of substrates such as gem-di-, trans-di-, cis-di, tri-, and tetra-substituted epoxides[31]. But for the typically *in vitro* purposes, the substrate is replaced by the trans-diphenylpropene oxide (t-DPPO), a radioactive aromatic epoxide and thus it will be used as a substrate in the present thesis. Recently, it was found that sEH plays an important role in the regulation of blood pressure and inflammation which makes it also a good target for the designing of drugs against several diseases, including high blood pressure, atherosclerosis, and kidney failure[32].

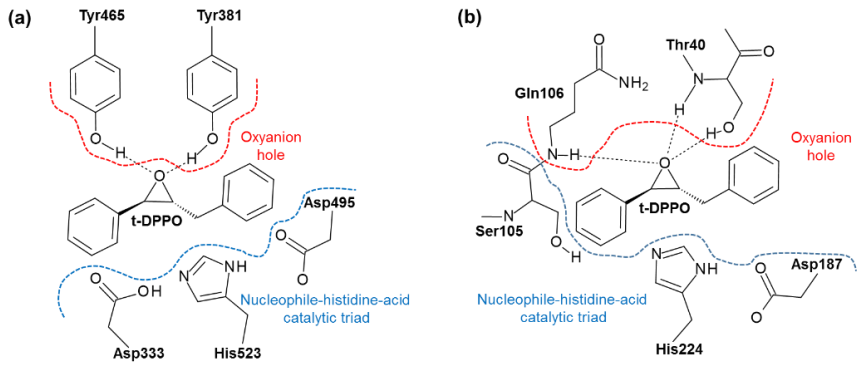
Both sEH and Cal-B belong to the family α/β hydrolases. The main feature of this family fold of proteins is the characteristic nucleophile-His-Acid catalytic triad evolved to efficiently operate on substrates with different chemical composition or physicochemical properties and in various biological contexts[33].

In case of sEH this triad is formed by Asp333, His523 and Asp495, as presented in Figure 4(a).Moreover an “oxyanion hole” is formed in the active sites by two tyrosine residues Tyr-381 and Tyr-465 which are assumed to stabilize the tetrahedral intermediate by protonation or hydrogen bonding of the oxygen atom of the epoxide[34]. Importance of these two residues for catalytic process was proven experimentally when 90% decrease in sEH activity was observed while one of tyrosine residues was mutated to phenylalanine[35].

What makes CAL-B so interesting now is that its active site is very close to the active site of the sEH (see Figure 4 (a)). As a serine hydrolase it has been proposed that catalysis proceeds by activation of the carbonyl of the substrate ester through an oxyanion hole made by two

residues (Thr40 and Gln106), which would stabilize the anionic tetrahedral intermediate, and the nucleophilic attack of a serine residue (see detail of active site in Scheme 1).11 This serine would be coupled to another two conserved residues thus forming a catalytic triad integrated by Ser105, Asp187 and His224, as can be seen in Figure 4 (b). The following figure depicts the similarities of the active site of both enzymes:

Figure 4: Active site of (a) sEH and (b) CAL-B enzymes together with bound (R,R)-trans-diphenylpropene oxide (t-DPPO) in active site.



As it was previously commented, and also can be seen in the above figure, the substrate of the reaction is the t-DPPO, an epoxide.

Epoxide-containing compounds are widely found in the environment from both natural and man-made sources and a large variety of aromatic and alkenic compounds are also metabolized to epoxides endogenously[36]. Epoxides, also known as oxiranes, are three-membered cyclic ethers that have specific reactivity patterns owing to the highly polarized oxygen-carbon bonds in addition to a highly strained ring[37]. Epoxides are highly important in organic synthesis because of their chemical versatility, as they readily react with almost every nucleophile, from halide to sulphur nucleophile through carbon nucleophiles, and also because they are an important source for producing chiral compounds.

Some reactive epoxides are responsible for electrophilic reactions with critical biological targets such as DNA and proteins, leading to mutagenic, toxic and carcinogenic effects[38].

1.2. - Computational Chemistry fundamentals

1.2.1. - Brief historical introduction[39]

Computational Chemistry is one of the most brand and newfangled branches of Chemistry. Computational Chemistry applies the methods of Theoretical Chemistry (quantum mechanics, statistical mechanics, thermodynamics, spectroscopy...) to study molecular systems such as biomolecules, polymers, and organic molecules for example.

The origin of Theoretical Chemistry can be pinpointed in 1925-1926, when two German physicists, Werner Heisenberg and Erwin Schrödinger proposed two different mathematical descriptions of quantum mechanics: Heisenberg proposed a matrix mechanics model in 1925 and Schrödinger a model based on differential equation (today known as the Schrödinger Equation). From this moment, the evolution of the quantum chemistry and consequently the

evolution of the computational chemistry[40] have reached a higher level of development. Quantum Chemistry developed rapidly during the second half of the 20th century and it was fuelled largely by breathtaking improvements in computer technology and, by 1998, when the Nobel Prize for Chemistry was awarded jointly to Walter Kohn and John Pople. Computational Chemistry increased in popularity and recently, the Nobel Prize for Chemistry 2013 was awarded jointly to Martin Karpus, Michael Levitt and Arieh Warshel "*for the development of multiscale models for complex chemical systems".*

1.2.2. - Computational Chemistry methods

1.2.2.1. - Quantum mechanics methods

1.2.2.1.1. - Schrödinger wave function

In agreement with the formulation of the Quantum Mechanics of Schrödinger, all the information of a system is contained in its wave function.

The deduction of the Schrödinger equation takes into count the Louis De Broglie hypothesis of the wave-particle duality and the Planck expression. With the appropriate mathematical operations[41], Schrödinger arrived to the mathematical equation which today is known as the Schrödinger equation dependant of the time (all the equations will be written only for one dimension, but can be extended to 3 dimensions, x- and y- and z-):

$$i\hbar\frac{\partial\Psi(x,t)}{\partial t} = \frac{\hbar^2}{2m}\frac{\partial\Psi^2(x,t)}{\partial x^2} + V(x)\Psi(x,t) \qquad (3)$$

For the case studied in the present thesis, the independent of the time wave function is enough. With the suitable mathematical operations[42] one can arrive to the following expression:

$$\hat{H}\Psi = E\Psi \qquad (4)$$

The molecular Hamiltonian operator is the combination of both the kinetic and potential operators for all the particles in the system which generally speaking, is built up of *n* nuclei and *e* electrons:

$$\hat{H} = \hat{T}_n + \hat{V}_{nn} + \hat{V}_{ne} + \hat{T}_e + \hat{V}_{ee} \qquad (5)$$

Where:

$\hat{T}_n$ is the kinetic energy operator of the nuclei

$\hat{T}_e$ is the kinetic energy operator of the electrons

$\hat{V}_{nn}$ is the potential energy between the nuclei

$\hat{V}_{ne}$ is the potential energy between nuclei and electrons

$\hat{V}_{ee}$ is the potential energy between the electrons

1.2.2.1.2. - Born-Oppenheimer Approximation.

The Born-Oppenheimer approximation was proposed by J. Oppenheimer and M. Born in 1927 and even nowadays it is a very useful and precise tool in molecular analysis. A molecule can be seen as an interaction between two types of particles: nuclei and electrons. Spectroscopic techniques carried out to study the behaviour of the molecules shown that the energy of a molecule can be described as a sum of the energies of its particles i.e. the energy of a molecule can be described as the sum of the energy of the nuclei plus the energy of the electrons As all scientists know, the mass of a nuclei is approximately 10^4 times more than the mass of an electron. Therefore, the nuclei movements are very slow in comparison to the electron movement. In another words, this situation can be seen as "frozen" nuclei interacting with an electronic gas, which is moving between the nuclei. In essence, the Born-Oppenheimer approximation allows us to calculate the energy of the nuclei separately from the energy of the electrons[43]. After these approximations, the equation (5) can be written as a new Hamiltonian, called electronic Hamiltonian, which mathematically is:

$$\hat{H} = \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} \qquad (6)$$

The equation 6 is best viewed as the movement of the electrons under the influence of an electric field generated by the nuclei.

1.2.2.1.3. - Ab initio methods.

Ab initio is a Latin expression which means 'from the beginning'. Therefore, *ab initio* computational methods[44] rely only on calculations based on theoretical principles without empirical data, usually the Schrödinger Equation. The *ab initio* methods give results with high accuracy, because it is more probable that one arrive to the true energy minimum. On the other hand, these types of methods have a high computational cost as well as a long time in some cases (depends directly of the studied system). With *ab initio* methods, circa 100 atoms can be treated. More will suppose a worthless computational effort.

1.2.2.1.4. - Semi- Empirical Methods.

Semi-empirical methods[44, 45] are based on *ab-initio* methods but they introduce some parameterisation in the equations, in order to simplify the calculations. The parameters introduced in the equations can be obtained from experimental observations or applying *ab-initio* methods.

The semi-empirical methods were developed because the *ab-initio* methods were too difficult and the computational expense was massive when one moved to more and more complex systems. Therefore, the main advantage of the semi-empirical methods is its speed as well as the number of atoms that can be treated, up to 10^3 atoms.

1.2.2.2. - Molecular Mechanics (MM) Methods.

Molecular Mechanics[46] (MM) were the first methods used in Computational Chemistry. They were mainly used for the study of systems in which basically one did not have to solve the Schrödinger Equation. Opposite to the Quantum Mechanical (QM) Methods, the Molecular Mechanics can be applied to study big systems, with lots of atoms, a thing which is impossible in a Quantum Mechanics due to the complexity of the resolution of the Schrödinger Equation.

Molecular Mechanics relies on classical approaches i.e. treats the atoms as hard spheres (no distinction between nuclei and electrons) and the bonding between atoms is treated as a classical spring, whose behaviour can be described with Hook's Law:

$$F = -k \cdot \partial x \qquad (7)$$

Where k is a constant characteristic for each bond and ∂x is the change in the position.

The interaction between the atoms is modelled by classical potential functions and the energy of the systems is calculated as the sum of the separate contributions: bonds, angles and dihedrals.

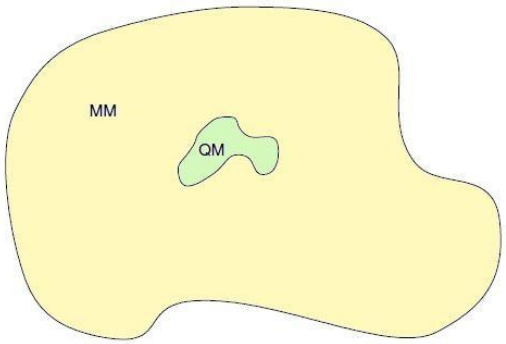
One can think that MM methods are no longer useful because of the arising of the QM Methods, but this is not true. Due to its speed and less computational expenses, MM methods are very popular in calculations of large systems as well as in molecular dynamics simulations (studies of the change in the state of a system with time). Some popular MM methods are OPLS-AA[47], AMBER[48] and CHARMM[49].

1.2.2.3. - QM/MM methods.

The arising of the Quantum Mechanics / Molecular Mechanics (QM/MM) Methods was due to the necessity of study systems in real situations i.e. in solution. In real systems, our solute is not in the vacuum, is actually surrounded by solvent molecules i.e. in a condensed phase.

QM/MM methods[50] were first proposed in 1976 by Warshel and Levitt[51] .The combination of both methods takes advantage of the best features of both: accuracy from QM methods and low computational effort from MM methods. The division of the systems by the QM/MM methods is shown in the following figure:

Figure 5: Division of the system in two subsystems done by QM/MM methods.[52]



QM/MM methods divide the studied system in two areas: one area, where the chemical process takes place (formation/breaking of bonds) QM methods are applied. In the other area there is no chemical reaction, but the processes of this area have a big influence on it, and it is treated with MM methods. These two subsystems are not isolated each other, there is a continue interaction between them.

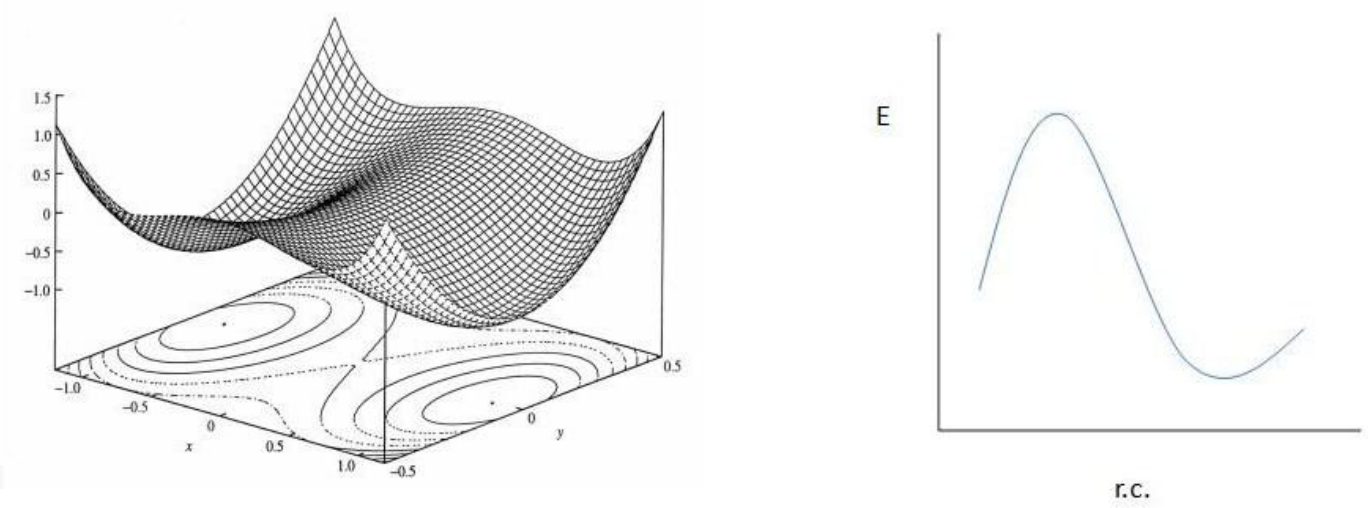
1.2.2.4 - Geometry optimization.

The geometry optimization is one of the most useful applications of Computational Chemistry. By performing geometry optimisation we pursue the objective of finding the lowest energy configuration of a molecule so performing a geometry optimization is in other words to perform an energy minimisation. This is very useful because molecular geometry is involved in lots of chemical and physical properties.

Geometry optimization is best viewed as a mathematical problem: '*How does one find a minimum in a multi-variable arbitrary function?*' This is the question which has to be answered. To do it, the chosen algorithm[53] selects one structure and calculated the energy of the system. If it is not the minimum energy conformation, the algorithm changes the structure and re-evaluates the energy of the molecule. The algorithm repeats this process until the minimum energy value is found.

Potential Energy Surfaces are based on the Born- Oppenheimer[43] approximation and can be defined as the set of values of the molecular potential energy corresponding to a set of possible relative position of its nuclei i.e. the relation between molecular structures and energies of these structures. The result is a Hypersurface of Potential Energy (HPE) which has 3N-6 dimensions, being N the number of atoms of the system but in practical cases HPE are not used because it is difficult to work with and obtain useful information from a 3N-6 (for example, if two non-linear tri-atomic molecules are being studied, the total number of atoms is 6 and hence the HPE have 12 dimensions). For this reason, normally 2D plots are used, choosing two representative parameters of the system (for example the bond length and the molecular angle) plus a third dimension, the change in potential energy (as can be seen in the Figure 6a). Despite the tree dimensional PES is the most common one, there are also 1-D PES, where only one representative parameter is chosen plus the energy (see Figure 6b). The following figure depicts two types of PES:

Figure 6: Examples of PES. On the right hand side, figure 5a, a 2D PES. On the left hand side, a 3D PES.



There are two main concepts to describe in a PES: the gradient and the Hessian.

The gradient is a vector composed of the first derivatives of the potential energy with respect to the positions of the atoms:

$$\overrightarrow{g_i} = \frac{\partial U}{\partial \overrightarrow{r_i}} \qquad \forall \, \mathrm{i} = 1, \ldots, \mathrm{N} \qquad (8)$$

Being N the number of atoms in the system.

The Hessian is a tensor, a square matrix which has as elements the second derivatives (or force constants) of the energy with respect to the displacement of two coordinates for all the atoms

$$H_{ij} = \frac{\partial^2 U}{\partial \overrightarrow{r_i} \partial \overrightarrow{r_j}} \qquad (9)$$

When the Hessian is diagonalized, the eigenvalues equation can be solved:

$$F\overrightarrow{u_i} = \lambda \vec{u} \qquad (10)$$

Where F is the diagonalized Hessian matrix, $\overrightarrow{u_i}$ is the vector that represents the curvature of the principal axes and represents the normal modes and λ are the eigenvalues.

As it can be seen in Figure 5a, there is one stationary point, which is characterized by a gradient equal to zero. Stationary points can be sorted into two groups:

i) Minima: There are minima when after the diagonalization of the Hessian matrix in the mentioned point give as a result nothing but positive values for the eigenvalues. This means that an infinitesimal displacement of the geometry of the system along the direction defined by any of the eigenvectors, $\overrightarrow{u_i}$, will lead to an increase in energy. This chemical meaning of minima is the presence of reactants, intermediates or products.

ii) First-order saddle points: There is a first-order saddle point when after the diagonalization of the Hessian matrix in the mentioned point, all the eigenvalues obtained are positive but for the first one, which is negative. This means that if an infinitesimal displacement is performed along the eigenvector corresponding to this negative eigenvalues, the energy will decrease. However, if the displacement is associated to an eigenvector with a positive eigenvalues, that will mean an increase in energy. The chemical meaning of a first-order saddle point is a transition state for the chemical reaction.

Another important element of a PES is the *Intrinsic Reaction Coordinate* or IRC[44, 50]. The IRC is the path that, starting from the TS leads to the reactants and products, following the direction of the gradient. In the initial point the transition structure has a gradient equal to zero. For this reason, the direction of the search must be previously defined for the transition vector. Mathematically, the IRC can be defined as the solution to the following differential equation:

$$d\vec{x} = \frac{\vec{g}(\vec{x})}{|\vec{g}(\vec{x})|}\, ds \qquad (11)$$

Being s is the path length from the transition structure, also known as the reaction coordinate.

2. –Aims

The present Bachelor's Degree Final Project (BDFG from now) has as a main goal the elucidation of the reaction mechanism of hydrolysis of t-DPPO in a cluster model employing theoretical methods. Moreover, further objectives are proposed:

i) Exploration of two possible reaction mechanisms.

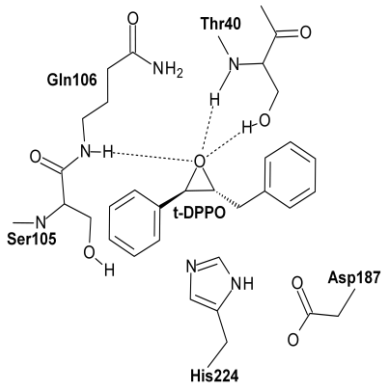
ii) Show the possible or not enantioselectivity of the enzyme.

iii) Propose mutations on wild-type CAL-B to increase its secondary activity as epoxidase.

3. - Theoretical Methods

The starting geometry was obtained from the Candida *Antarctica Lipase B* (CAL-B) PDB entry 1TCA[54] which consists of a monomer. PROPKA3 semi empirical program[55] was used to recalculate the pKa values of the titratable protein residues to verify their protonation states at a pH of 7. According to the results, most residues were found at their standard protonation state, except for the Asp187 residue that should be protonated and His224 which was protonated in δ-position. Additionally, the disulphate bridges between Cys22 and Cys64, Cys216 and Cys258, and Cys293 and Cys311 were defined. Since the total charge of the system was neutral, no counterions were required, so only the hydrogen atoms are introduced.
Once this was done, a cluster model is built up, in order to mimic the active site of the native enzyme i.e. bearing in mind the conservation of the catalytic triad and the oxyanion hole found in the active site of the native enzyme. Henceforth, the cluster model will be formed by the inhibitor, which will be modified for the t-DPPO, for the reasons described in the Introduction, and the residues Asp 187, Ser105, Gln106, Thr40 and His224. To reduce the computational effort, all residues were saturated in Cα positions.

Figure 7: Designed cluster model.



For the study of the Mechanism A, a water molecule, is introduced in the cluster model by means of the GaussView 5.0 program[56].After that, two QM optimizations were carried out:
First optimization: The oxyanion distances are constrained by the application of a force constant of 2500 kJ·mol^{-1}·Å^{-2}:

d (H^{Gln106}- O^{epox}) =2.66 Å

d (N-H^{Thr40}- O^{epox}) =3.14 Å

d (O-H^{Thr40}- O^{epox}) =2.15 Å

Second optimization: The system was all free optimized i.e. no constraints were applied. After the optimization, the variation of the distances was not so large so were sorted as right:

d (H^{Gln106}- O^{epox}) =2.61 Å

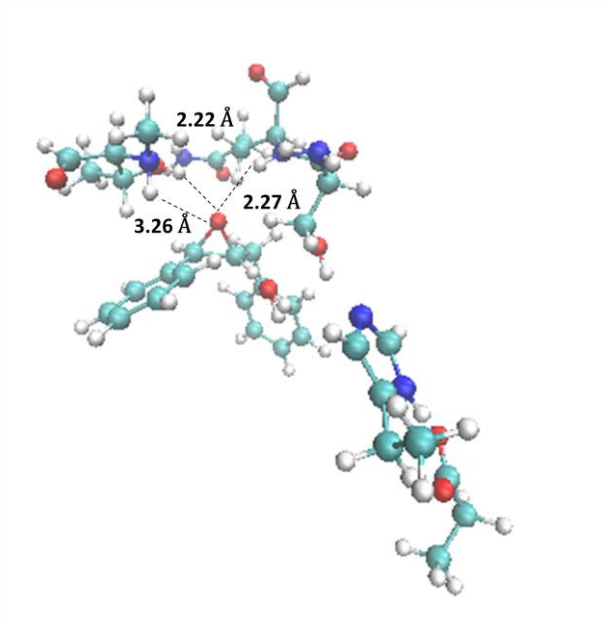
d (N-H^{Thr40}- O^{epox}) =3.12 Å

d (O-H^{Thr40}- O^{epox}) =2.19 Å

All the optimizations were performed by means of the AM1 semi empiric Hamiltonian[57]. The rest of the system (protein plus water molecules) were described using the OPLS-AA and TIP3P[58] force fields, respectively, as implemented in the fDYNAMO library[59].

The following Figure shows the cluster model, ready to work with:

Figure 8: Starting optimized cluster model for mechanism A.



For the study of Mechanism B, a different treatment was applied. First of all, the Ser105Asp mutation is performed i.e. the Asp must be introduced in the system, with elimination of the Ser residue. This mutation is performed by means of the Accelerys Discovery Studio 3.5 program[60].PROPKA3 semi empirical program[55] was used to recalculate the pKa values of the titratable protein residues to verify their protonation states at a pH of 7. According to the results, most residues were found at their standard protonation state, except for the Asp187 residue that should be protonated and His224 which was protonated in δ-position exactly as before but now, another residue, the Asp105 must be protonated as well.

The following table summarizes the results of the PROPKA3 checking:

Table 1: A calculation of pKa of titratable residues in the protein environment by means of PROPKA3.

Residue	pKa	pKa in the protein environment
Asp 105	11.41	3.80
His 224	5.19	6.5
Asp 187	2,80	3.80

The same disulphate bridges as before were defined as well. Now, the inhibitor is changed to the t-DPPO as before.

Once this is done, the enzyme which is introduced is solvated by its introduction in a orthorhombic box of water molecules of 100.921 x 79.5 x 79.5 Å of side. After that, a series of QM/MM optimizations are performed:

First optimization: The oxyanion distances are constrained by the application of a force constant of 2500 $kJ \cdot mol^{-1} \cdot Å^{-2}$:

d (H^{Gln106}- O^{epox}) =2.26 Å

d (N-H^{Thr40}- O^{epox}) =2.24 Å

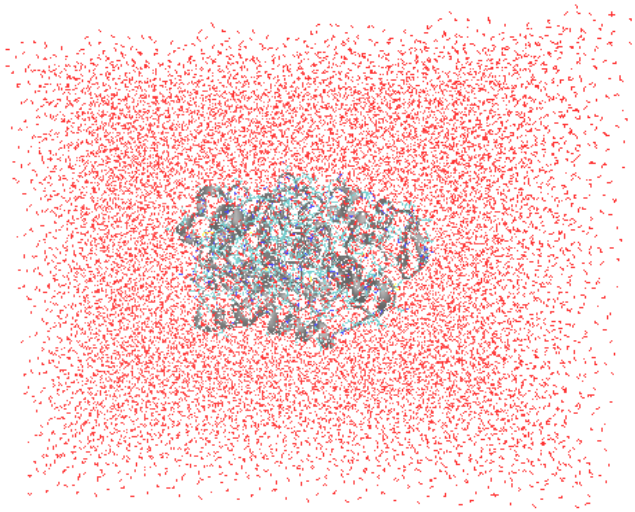
d (O-H^{Thr40}- O^{epox}) =2.13 Å

Second optimization: The system was all free optimized i.e. no constraints were applied. After the optimization, the distances had the exact value as defined by the constraint.

To reduce the computational effort of the optimizations, Atoms belonging to molecules found at a distance less or equal to 25 Å of the oxygen of the epoxide were defined as flexible. The rest of atoms are kept frozen and only the substrate of the reaction was defined as QM. The rest was treated as MM with non-bonding interactions treated by means of the application of cut-offs, in particular, using a switching scheme with the radius ranging from 12 to 15 Å.

The following figure shows the solvated system:

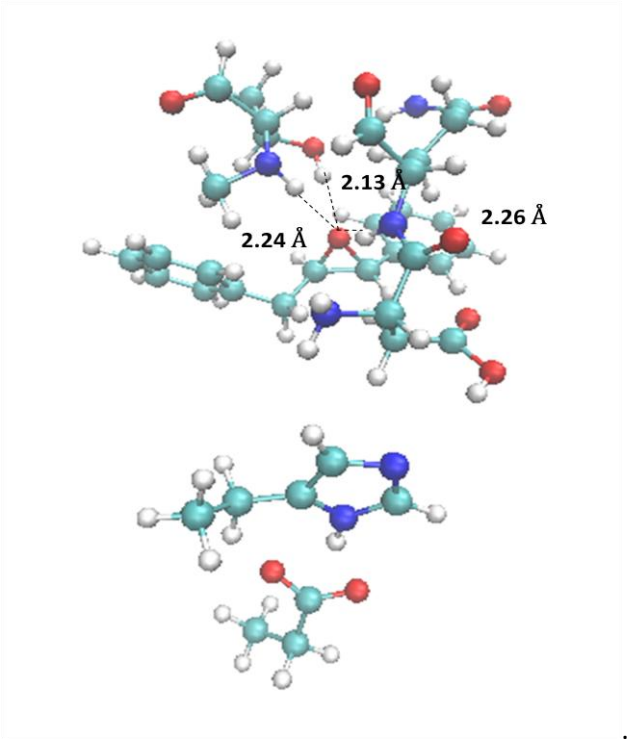
Figure 9: Solvated system.



After that, the residues and substrate that will form the cluster model are selected and a QM optimization is performed with the oxyanion distances constrained at the values of the QM/MM optimization. After the optimization, a new one is performed with all free i.e. without constraints.

The following Figure depicts the cluster model ready to work with:

Figure 10: Starting optimized cluster model for mechanism B.



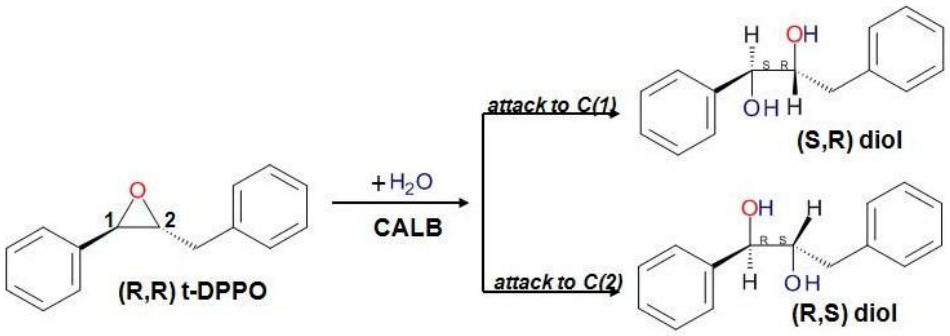
The different steps of every proposed mechanism are explored by means of 2D or 2D Potential Energy Surfaces. The stationary points on the different PES are located by using the Baker search algorithm with a step number equal to 1000 and a step of 0.1 Å. At any time, the criteria of convergence is 1 kJ·Å^{-1}. After localizing the stationary points, frequency calculations were carried out to verify that the structures represent true minima or first-order saddle points on the gas phase PESs. Once first-order saddle points were located and characterized, the Intrinsic Reaction Coordinate (IRC) path was traced down from the saddle points to the corresponding minima using the full gradient vector. The global r.m.s. residual gradient in the optimized structures was always less than 0.04 kcal mol^{-1} Å^{-1}.

All the energy values shown in the potential energy surfaces, as well as in the energy profiles, are in kcal/mol units.

4.- Results

As the substrate of the reaction is not symmetric, its hydrolysis mat take place on C1 or C2 of the epoxide, leading to different enantiomers, as showed in the following figure:

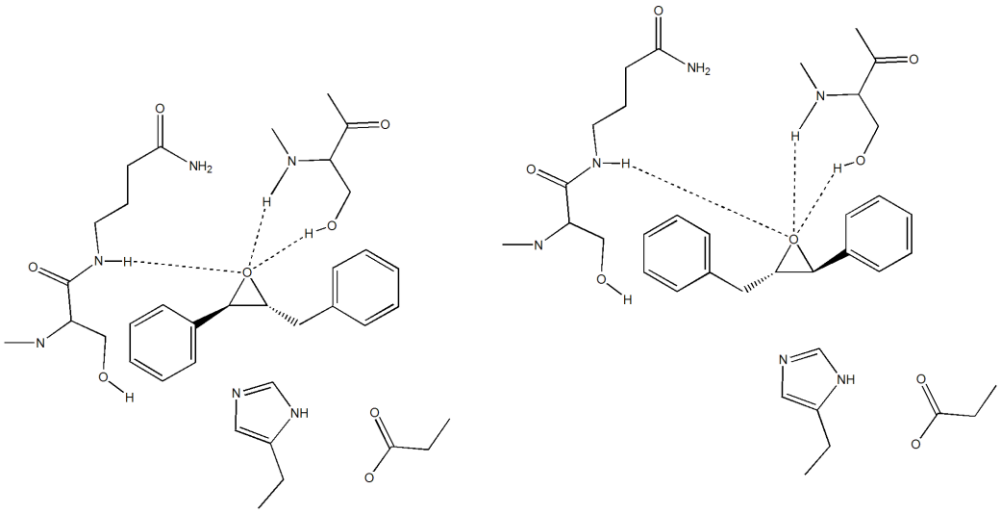
Scheme 4: Schematic representation of the possible enantiomeric products of the hydrolysis of (R,R) t-DPPO



The substrate of the reaction is the t-DPPO, known in this BDFG as Conformation 1, but the epoxide can also exist in its other enantiomer, the c-DPPO, known as Conformation 2.

The following figure depicts both conformations for the substrate:

Figure 11: Schematic representation of the two possible conformations of the epoxide

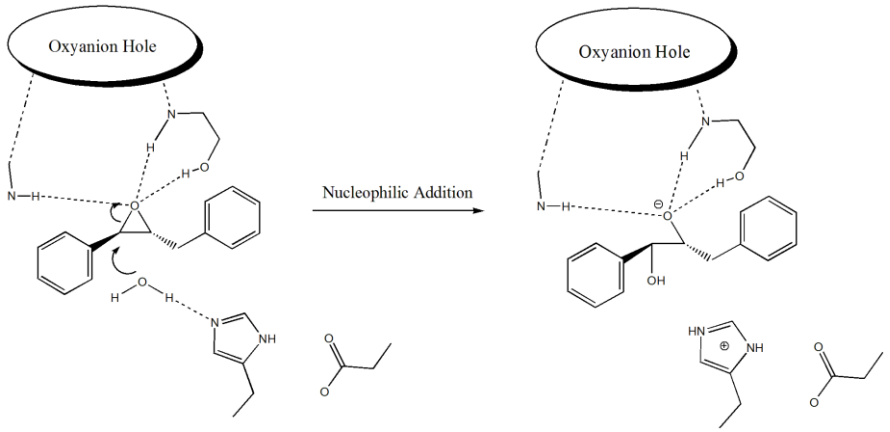


Conformation 1: t-DPPO **Conformation 2: c-DPPO**

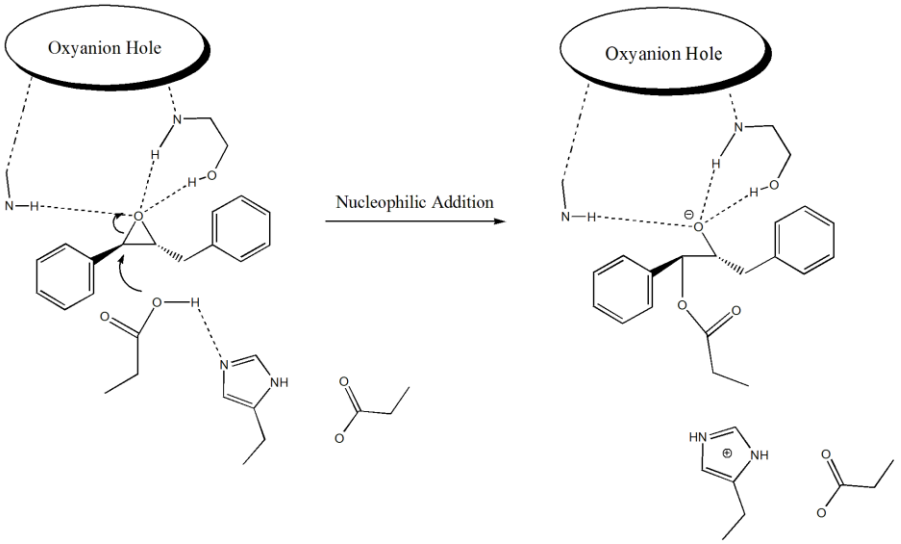
Two possible scenarios are going to be explored. Mechanism A is a nucleophilic direct attack of water whilst Mechanism B is a nucleophilic attack of the residue of the enzyme, as shown in the following figure:

Scheme 5: Schematic representation of the two possible mechanisms of hydrolysis

A:



B:



Hence, in this BDFG, the hydrolysis to both C1 and C2 of the t-DPPO for Mechanism A are presented, just for the t-DPPO (Conformation 1).

Mechanism B will be presented for both wild type CAL-B and for the mutated enzyme, the Ser105Asp CAL-B, for both conformations.

4.1. - Study of Mechanism A.

The strategy that had been used to study the Mechanism A is the exploration of the total reaction in two steps. In the first step, a 3D PES is explored by the use of two antisymmetric combination of distances in both the X and Y axes.

In the X axe the antisymmetric combination that had been employed is:

$$r1 = d(O^{epox} - C_i) - d\,(C_i - O^{wat}) \qquad (12)$$

i may take as values 1 or 2, depending on the carbon that is being involved in the reaction.

In the Y axe the antisymmetric combination that had been employed is:

$$r2 = d(O^{wat} - H^{wat}) - d\,(H^{wat} - N^{His}) \qquad (13)$$

This will start from the reactant complex (RC) of the reaction and will lead to the Intermediate 1 (INT1).

In a second step, a 2D PES will be explored, again, by employing only one antisymmetric combination of distances. This antisymmetric combination of distances is described in the following equation:

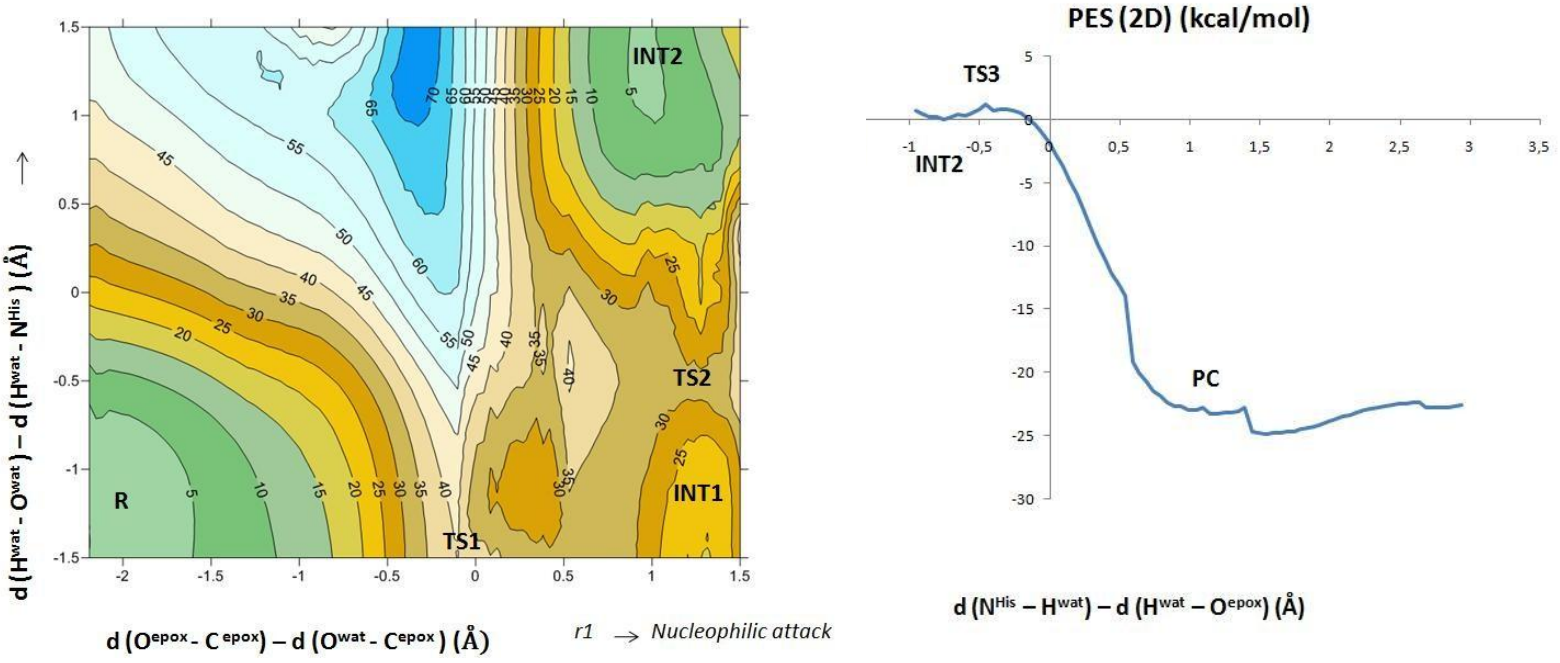
$$r1 = d\left(N^{His} - H^{wat}\right) - d\left(O^{epox} - H^{wat}\right) \qquad (14)$$

Equations 11 to 13 will be particularized to each starting structure.

i) Attack on C1.

The following figure shows the 3D PES obtained for the first step of the study of the nucleophilic attack to C1, mechanism A and the corresponding 2D PES.

Figure 12: On the left hand a 3D PES corresponding to formation of the C1-O^{wat} bond and transference of the H of the water to the Nε of the His 224. On the right hand side, 2D PES corresponding to the transference of the H of the water to the O of the epoxide.



During the calculation of the PES, to prevent the breaking of the C1 – C2 bond of the epoxide, a constraint between the mentioned atoms had been applied. The bond is constrained to a distance of 1.49 Å, by means of a force constant of 5000 kcal·mol^{-1}· Å^{-2}.

The PES had been generated with a step length of 0.1 Å for the nucleophilic attack coordinate (X- Axe) and 0.05 Å for the transference of the hydrogen (Y-Axe).

In order to constrict the movement of the atoms that define the reaction coordinates, a force constant of 2500 kcal·mol^{-1}· Å^{-2} had been applied.

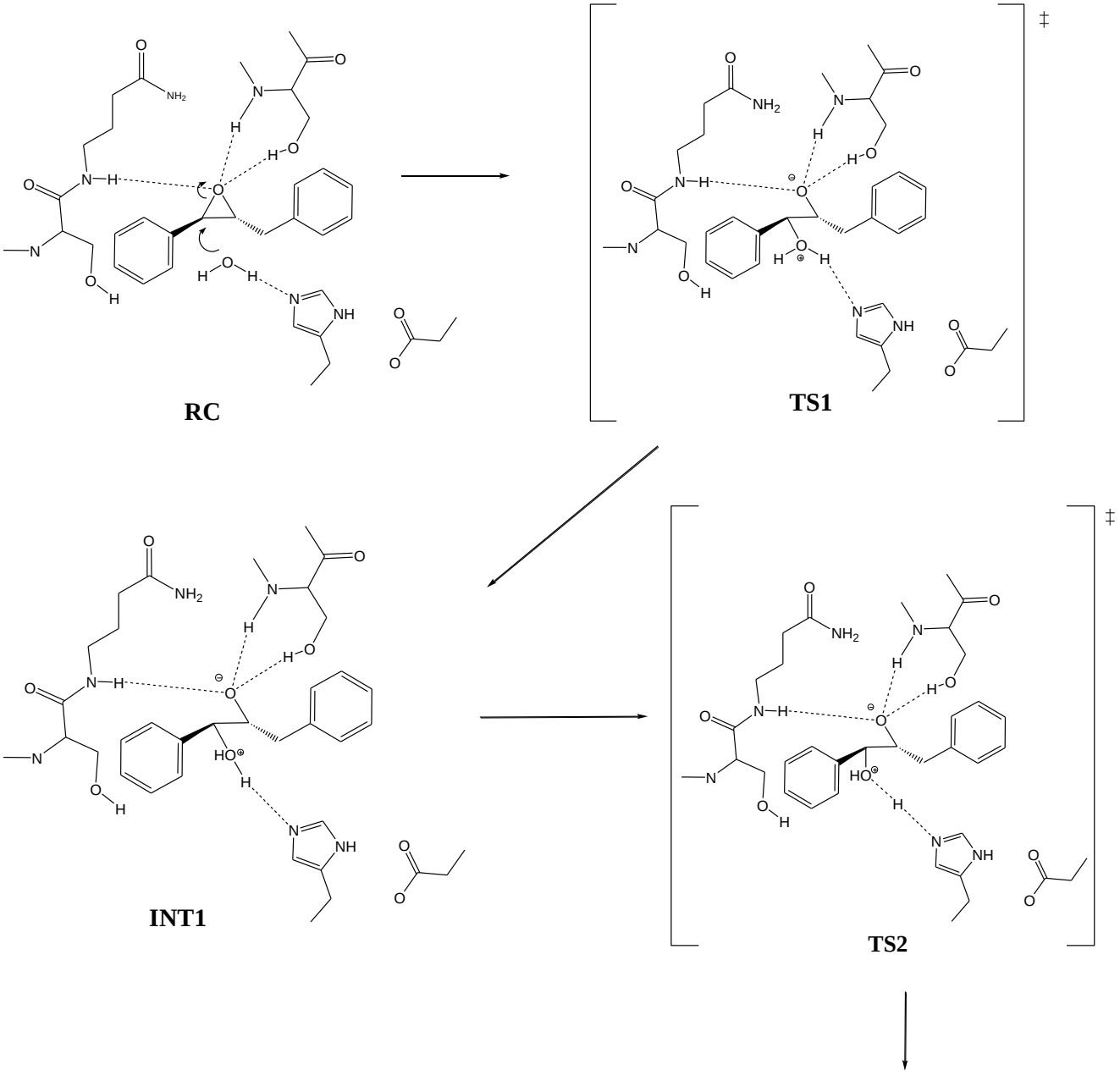
After the calculation of every single point, an optimization of the system with the *lbfgsb* during 5000 steps had been performed, applying a criteria for convergence of 1 kJ·Å^{-1}.

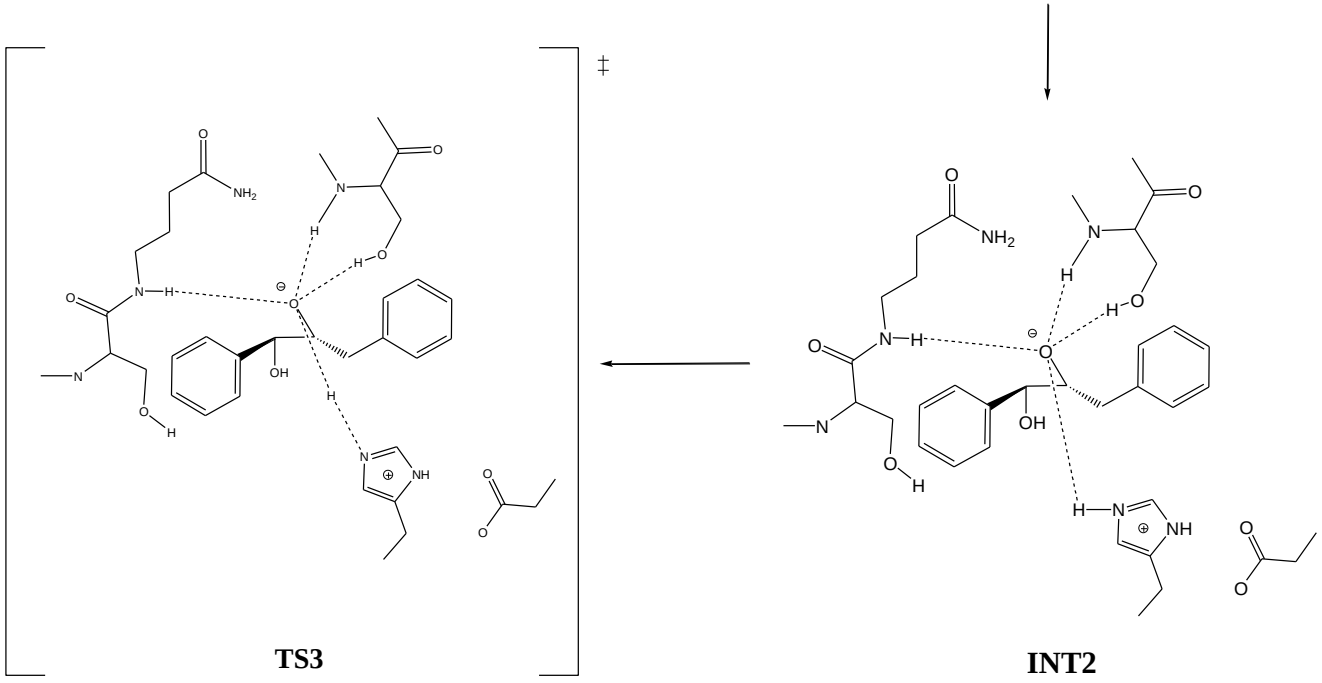
As we can see by simple inspection of the 3D PES, this first step will go in two steps so the mechanism is stepwise, with two transition states and two intermediates, so the products of this PES will be the INT2.

The 2D PES has the INT1 as 'reactants', and goes through one TS, TS3, to products, PC.

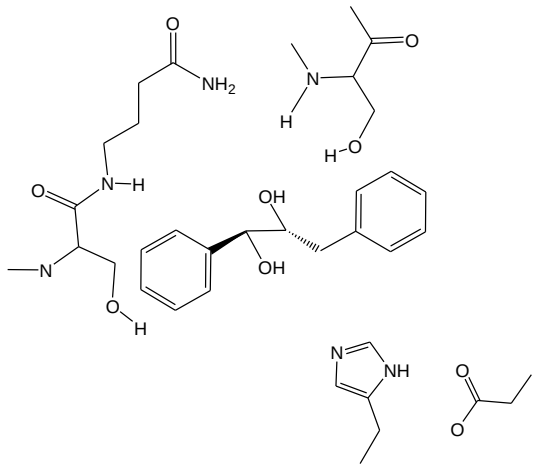
The following scheme shows the overall mechanism of the reaction:

Scheme 6. Stepwise mechanism of the hydrolysis of the t-DPPO by nucleophilic attack of water catalyzed by CAL-B.





TS3 **INT2**



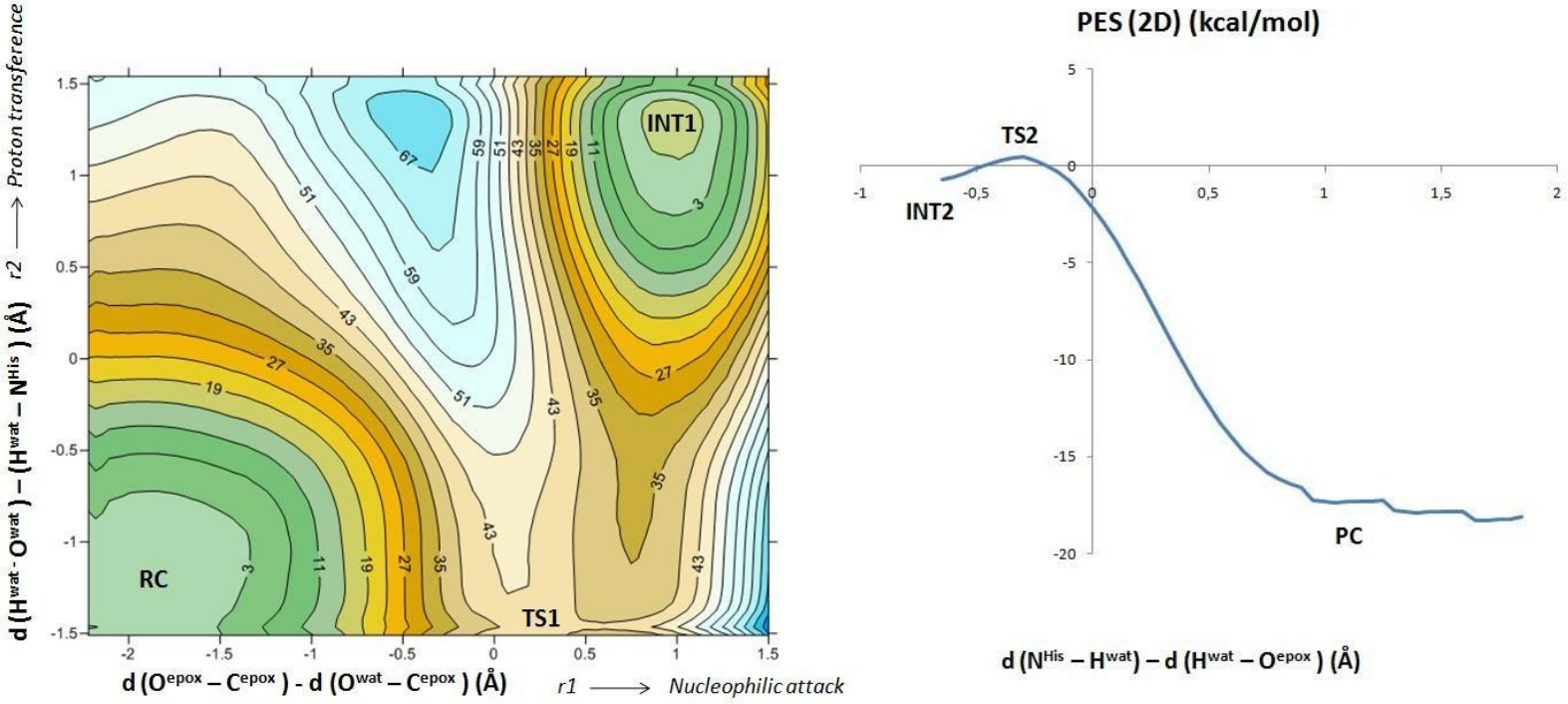
PC

ii) Nucleophilic attack on C2.

- *Nucleophilic attack on C2 and transfer of the H from the water to the Histidine.*

The following figure shows the 3D PES obtained for the first step of the study of the nucleophilic attack to C1, mechanism A.

Figure 13: 3D PES corresponding to formation of the C2-O^{wat} bond and transference of the H of the water to the Nε of the His 224 and *2D* PES corresponding to the transference of the H of the water to the O of the epoxide.



As we can see in Figure 16, this time the reaction only shows one TS, TS1, and goes directly to products, which in this case is INT1. Hence, the mechanism is concerted. However, when the IRC analysis is carried out, we obtain different results, as the formation of an intermediate like the one formed in the attack on C1 is observed. The reason for not to be observed in the PES is that in the PES, only potential energy is taken in count. Sometimes, the system adopts lower energy conformations which we cannot see in the Potential Energy Surface.

To obtain more realistic surfaces, Molecular Dynamics are used, like calculation of Potential Mean Forces (PMF). These methods rely on statistic methods like WHAM[62] (weighed histogram analysis methods) combined with a tool to explore the different values of the reaction coordinate along the reaction path, like umbrella sampling methods. Although these methods can help to a better understanding of the PES of the Figure 16, they are beyond the scope of this work. Hence, no Molecular Dynamics are shown here.

Despite the difference in the first step of the reaction between the nucleophilic attack to C1 and the attack to C2, the second step, the transference of the proton from the Histidine 224 to the O of the epoxide, the step that leads to products, it is exactly the same for both, as can be seen if Figure 17 is compared to Figure 15.

Then, the mechanism for the hydrolysis on C2 is exactly the same as the one for C1. For limited space of the BDFG and because of the similarity of the mechanism with the one shown in Figure 16, the mechanism for C2 will not be shown. One can just observe Figure 16 and imagine the mechanism for C2 just by visualizing the nucleophilic attack on C2 instead on C1.

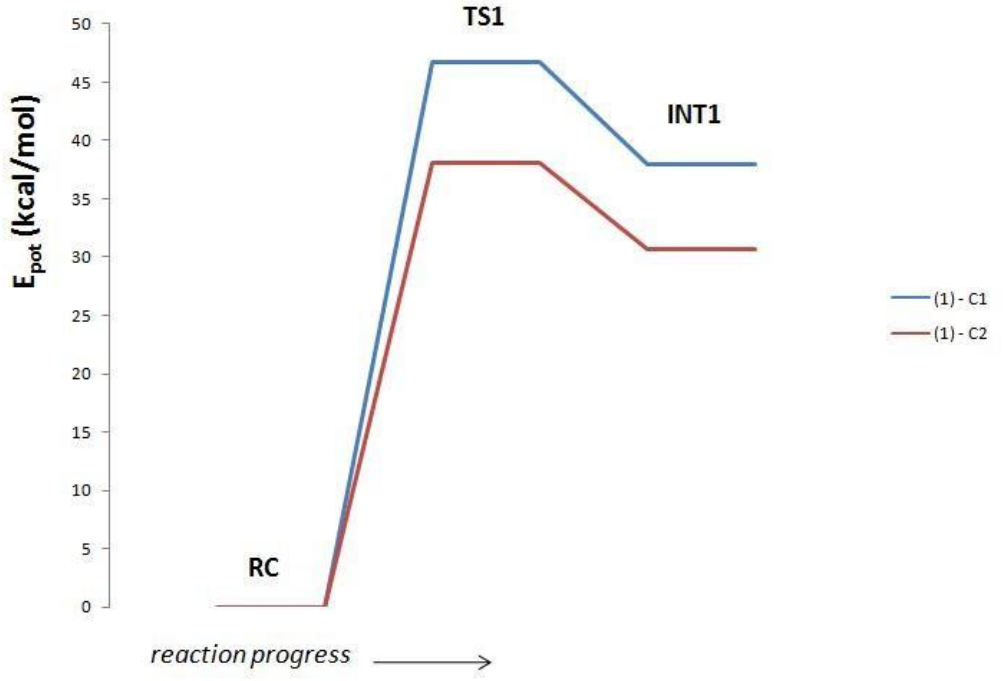
The following table summarizes the energetic values of Potential Energy for all the structures obtained for the hydrolysis of the t-DPPO (Conformation 1) for both C1 and C2.

Table 2: Potential Energy values (in kcal/mol) for the structures of the mechanism of hydrolysis of t-DPPO (Conformation 1), Mechanism A

Structure	*C1*	*C2*
RC	0.00	0.00
TS1	46.76	38.1
INT1	37.95	30.7

The results of the Table 1 can be plotted out into a Potential Energy profile in which it is easier to see the difference in terms of energy:

Figure 14: Energy profile for the nucleophilic attack of the water to C1 and C2 of t-DPPO (Conformation 1).



4.2. - Study of Mechanism B on Ser105Asp Mutated CAL-B.

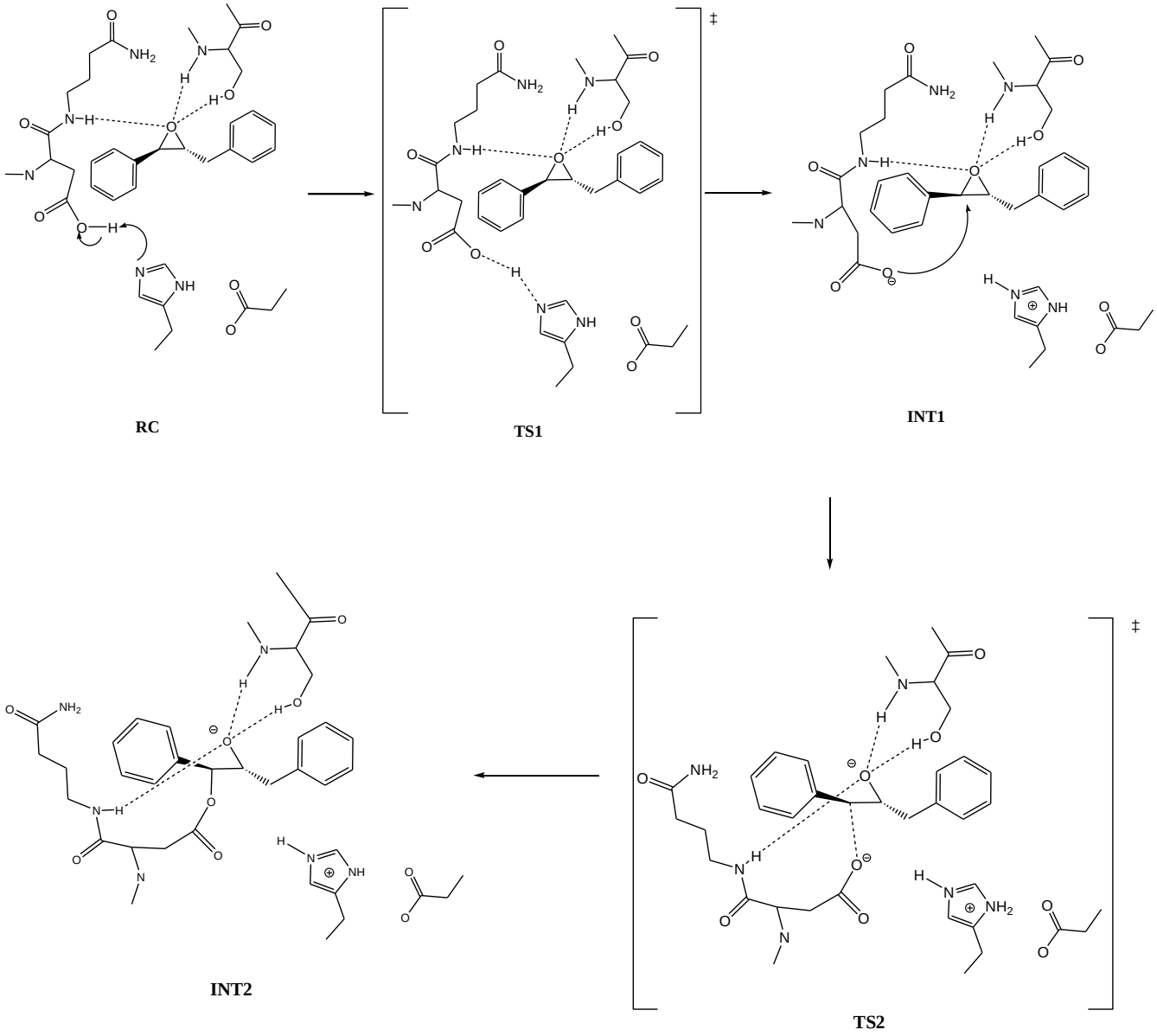
As mentioned in the introduction, the active site of CAL-B presents noticeable similarities with the active site of sEH. However, in the wild type CAL-B the role of the nucleophile in its catalytic triad is played by a serine (Ser105) instead of an aspartate (Asp333) residue. In fact, as had been shown experimentally, the Asp333Ser variant of sEH loses its catalytic activity. Thus, it can be concluded that wild type CAL-B should present the same behaviour. However, Ser105Asp mutated CAL-B should perfectly mimic the active site of wild type of sEH. Bearing in mind these similarities, apart from differences in the rest of scaffold of proteins, the main difference in active site for mutated CAL-B is the formation of an oxyanion hole; the reaction should proceed through a similar path as for sEH. Furthermore, as we already have

demonstrated, the role of oxyanion hole in sEH is to stabilize the negatively charged oxygen of observed intermediate, because other possible role of being donor of proton has been already excluded.

The mechanism that is going to be explored is a multistep mechanism whose first step, which will be named B(1) is the nucleophilic attack of the residue of the enzyme (Ser105 or the Asp105 after the mutation). Then, it can proceed by two different pathways, which will be called B(2) and B(3).

B(1) was experimentally demonstrated by Mulholand *et al*[63] and is depicted in the following scheme:

Scheme 7. First step of the multistep mechanism B, named B(1), studied experimentally by Mulholand et al.



This step had been studied by means of theoretical methods for both C1 and C2 in this BDFG for both Conformation 1 and Conformation 2 in wild type CAL-B and in Ser105Asp mutation.

-Ser105Asp Mutation. Study of the first step of Mechanism B, B(1).

Firstly, the results for Conformation 1 will be shown. The strategy that had been followed is the exploration of the step by means of two Potential Energy Surfaces: firstly, one 3D PES will be explored, that will lead us to the second intermediate of the reaction. To make things easier, to locate the second transition state, and hence to arrive to the second intermediate, a 2D PES surface is carried out, starting from Intermediate 1.

For the 3D PES, the following equations had been employed:

X-Axe: normal combination of distances:

$$r1 = d(O^{Asp} - C_i) \qquad (15)$$

Y-Axe: Antisymmetric combination of distances:

$$r2 = d(O^{Asp} - H^{Asp}) - d\,(H^{Asp} - N^{His}) \qquad (16)$$

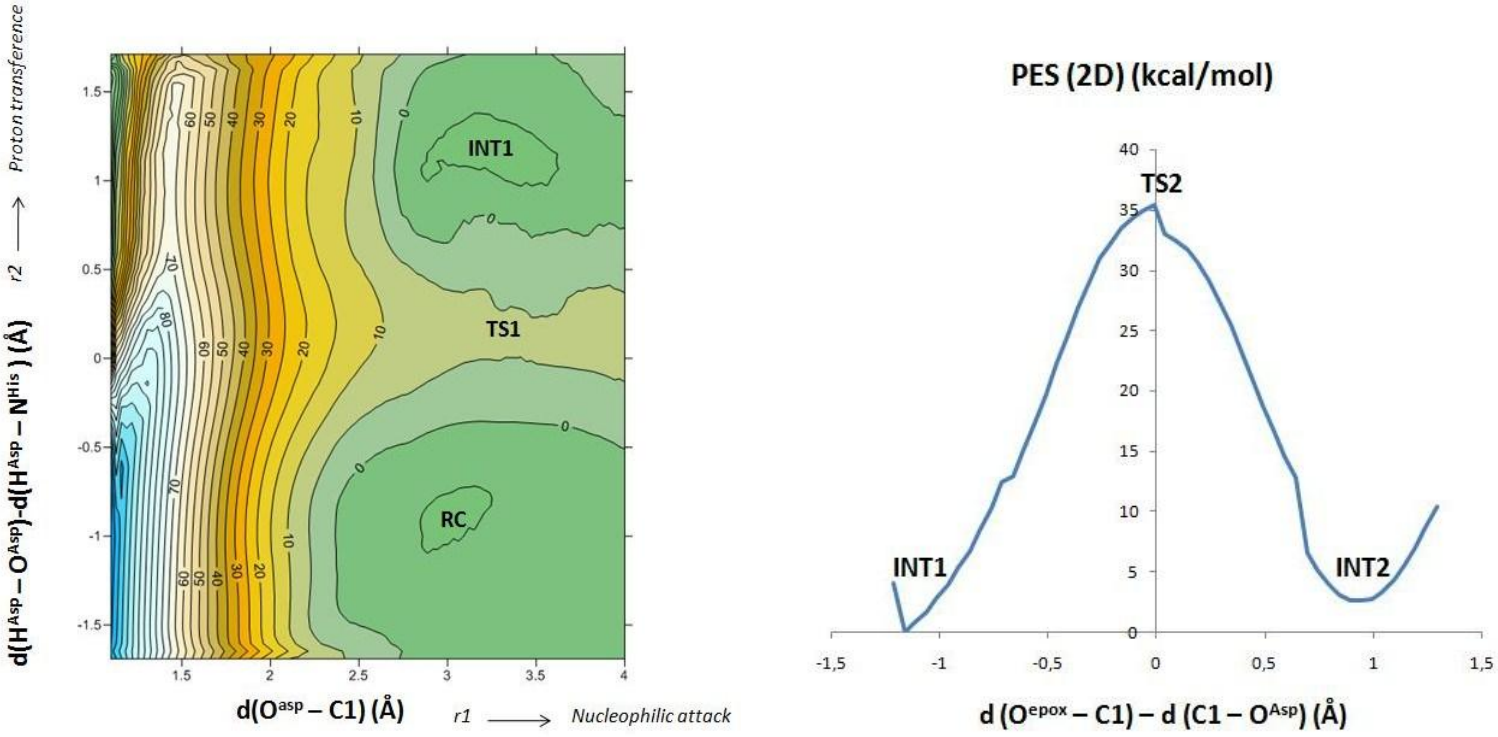
While for the 2D PES, the employed equation is the following antisymmetric combination of distances:

$$r1 = d(O^{epox} - C_i) - d\,(C_i - O^{Asp}) \qquad (17)$$

- *Conformation 1.*

The following figure shows the surfaces that have been explored:

Figure 15: 3D PES corresponding to formation of the C1-O^{Asp} bond and transference of the H of 105Asp to the Nε of the His 224 and assistant *2*D PES.



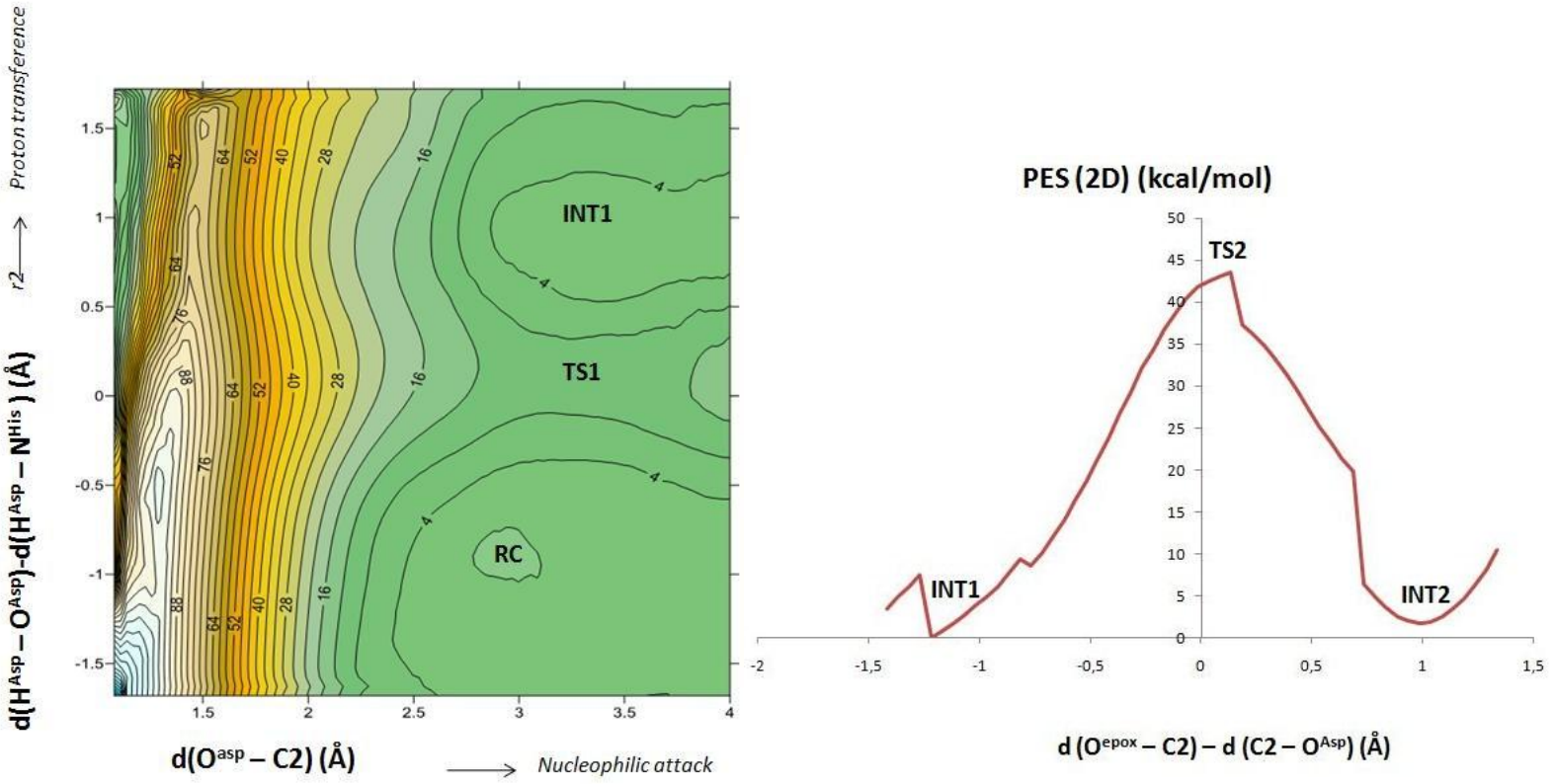
As we can see, the PES shows a stepwise mechanism via formation of two transition states. The first one is associated to the transference of the proton from the Asp105 to the His224 and the second, to the formation of the C1-O^{Asp} bond.

For a better localization of the second transition state, as described at the beginning of the section, a 2D PES surface is carried out, showed on the right hand side of Figure 18.

i) Attack on C2.

The following figure shows both the 3D and 2D potential energy surfaces for the exploration of this step:

Figure 16: 3D PES corresponding to formation of the C2-O^{Asp} bond and transference of the H of 105Asp to the Nε of the His 224 and assistant *2*D PES.



- ***Conformation 2.***

For the study of Conformation 2, a different strategy had been taken. As we already know how the reaction will behave, it is not necessary to make 3D PES. Hence, for the study of Conformation 2, only 2D PES had been used.

For the localization of the first transition state, the one associated to the transference of the proton from the Asp105 to the Nε of the His 224, a 2D PES with an antisymmetric combination of distances had been used. The equation is:

$$r1 = d(H^{Asp} - O^{Asp}) - d\,(H^{Asp} - N^{His}) \qquad (18)$$

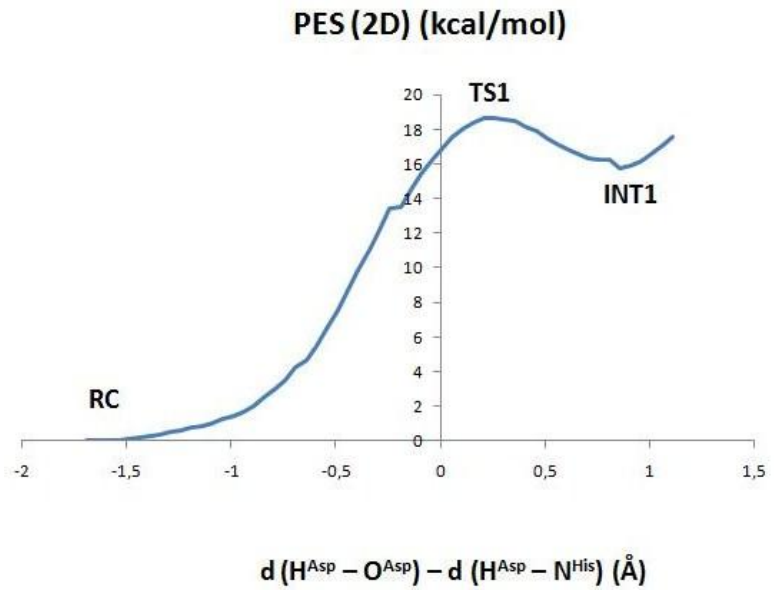
For the localization of the second transition state, the formation of the $C^{epoxide}$-O^{Asp} bond, another antisymmetric combination of distances had been employed, which is:

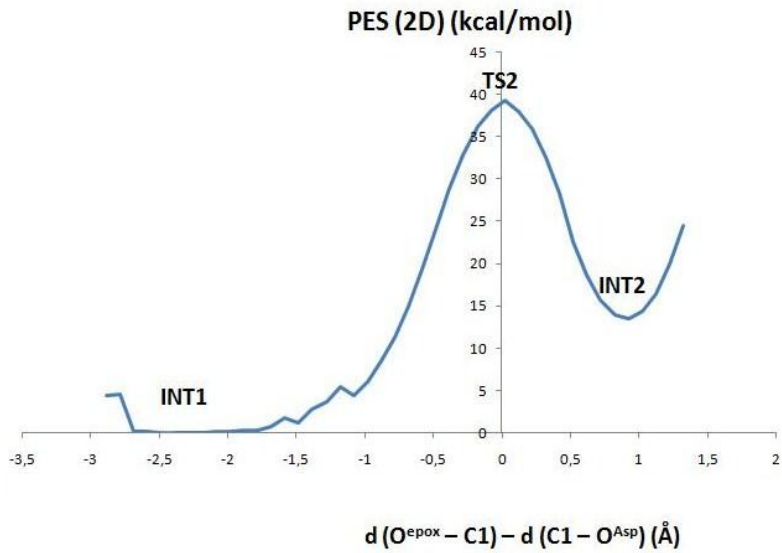
$$r1 = d(O^{epox} - C_i) - d\,(C_i - O^{Asp}) \qquad (19)$$

i) Attack on C1

The potential energy surfaces that had been used for the study are showed in the following figure:

Figure 17: 2D PES corresponding to the transference of the proton from the Asp105 to the Nε of the His 224 (left hand side) and 2D PES corresponding to the formation of the C1-O^{Asp} bond (right hand side).



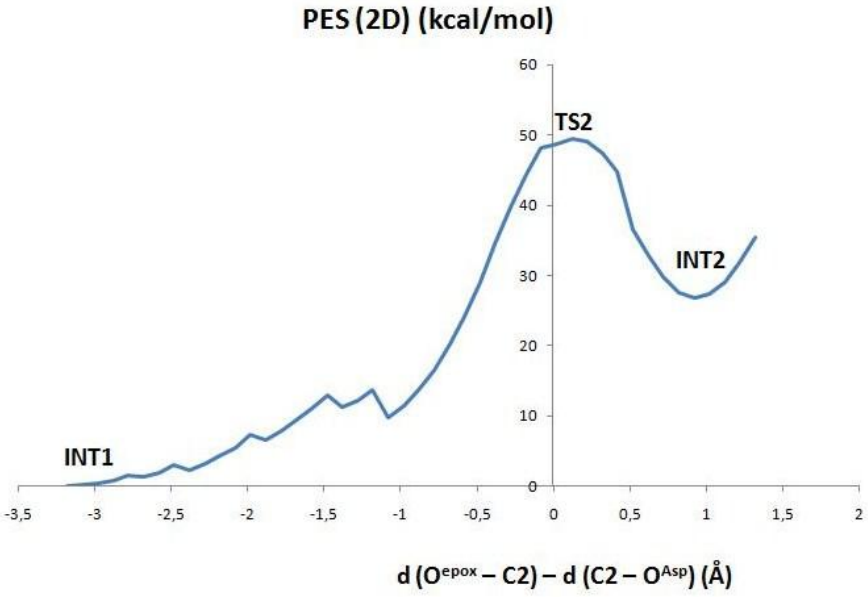


ii) Attack on C2.

For the attack on C2, only the PES with the TS related to the formation of the C2-O^{Asp} bond is showed. This is because the first step, the transfer of the proton, is the same for both C1 and C2, as neither of them is taking part in the reaction.

Henceforth, in the following figure, only the previously mentioned PES is presented:

Figure 18: 2D PES employed to study the nucleophilic attack to C2



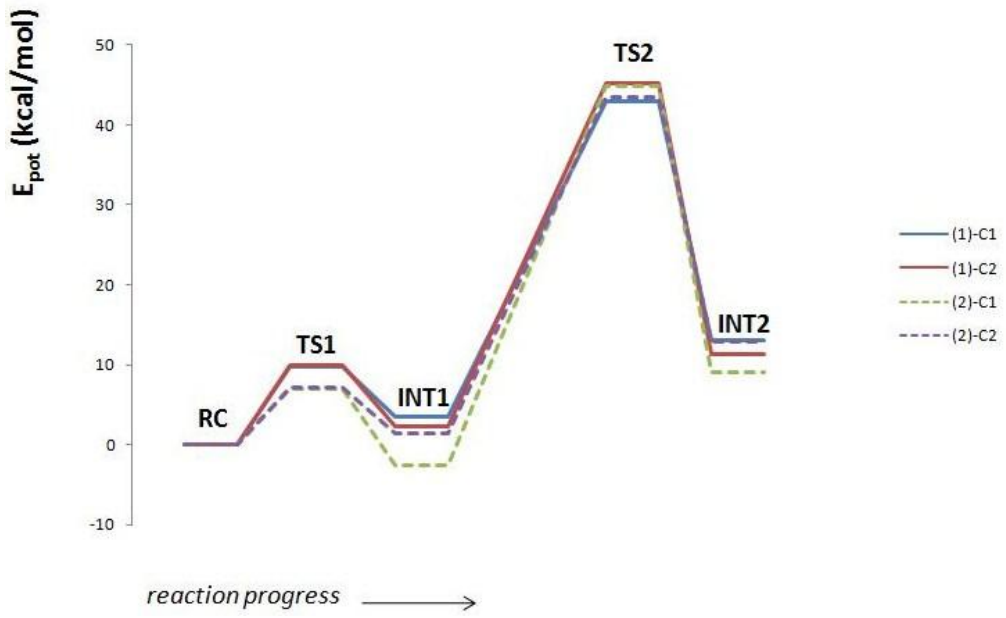
The following table contains a summary of the Potential energy values for each structure:

Table 3: Potential Energy values (in kcal/mol) for the structures of first B(1) for both conformations: (1) = Conformation 1; (2) = Conformation 2.

	(1)-DPPO C1	*(1)-DPPO C2*	*(2)-DPPO C1*	*(2)-DPPO C2*
RC	0.0	0.00	0.00	0.00
TS1	9.8	9.85	7.05	7.05
I1	2.6	2.56	-0.47	-0.47
I1	-0.8	0.34	2.05	-1.78
TS2	39.5	42.96	44.84	42.15
I2	9.6	9.13	11.61	11.61

The data of Table 3 can be plotted out into a Potential Energy profile, as we can see in the next figure:

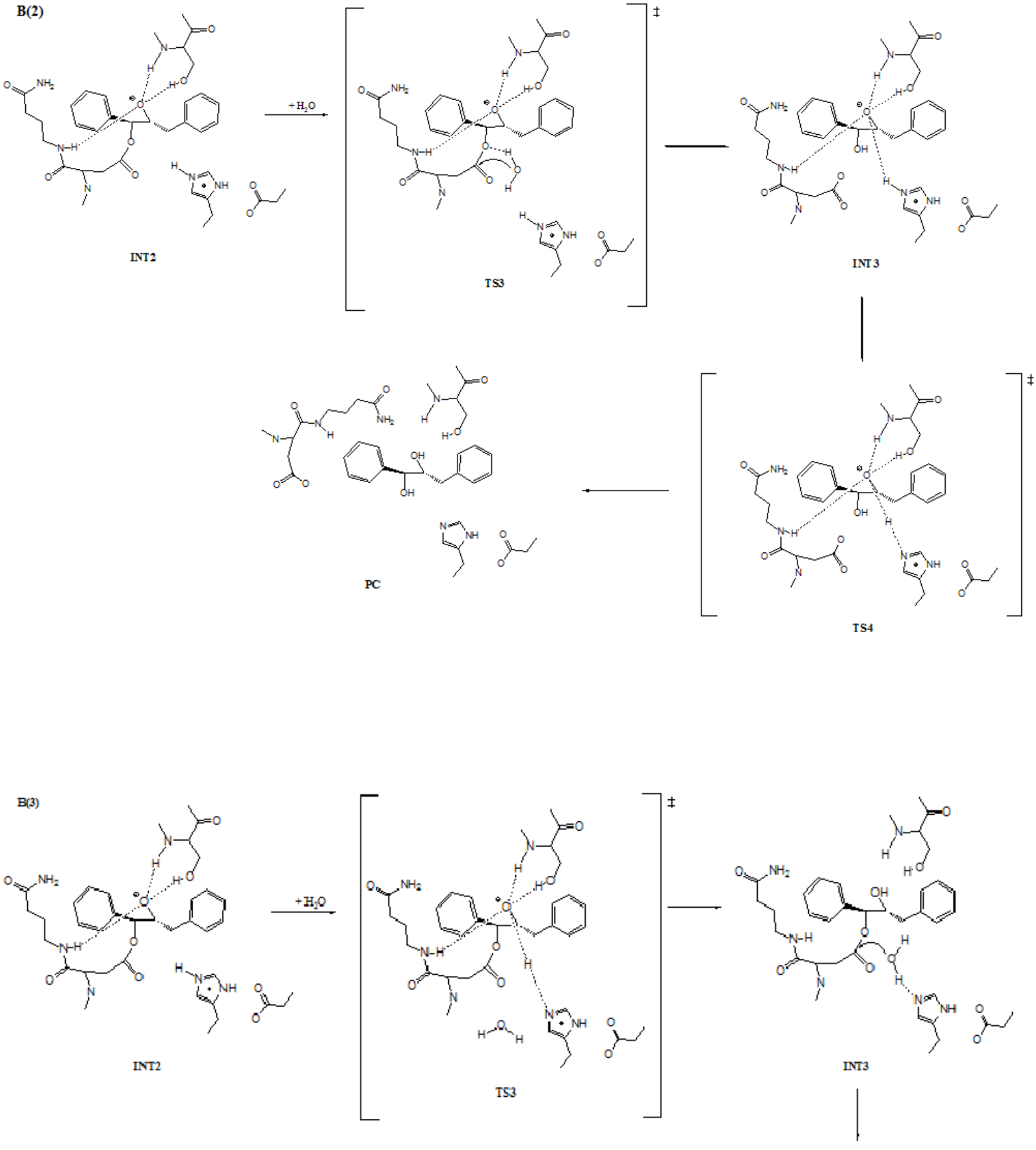
Figure 19: Energy profile for the first step of Mechanism B to C1 and C2 of the two conformations of the substrate.

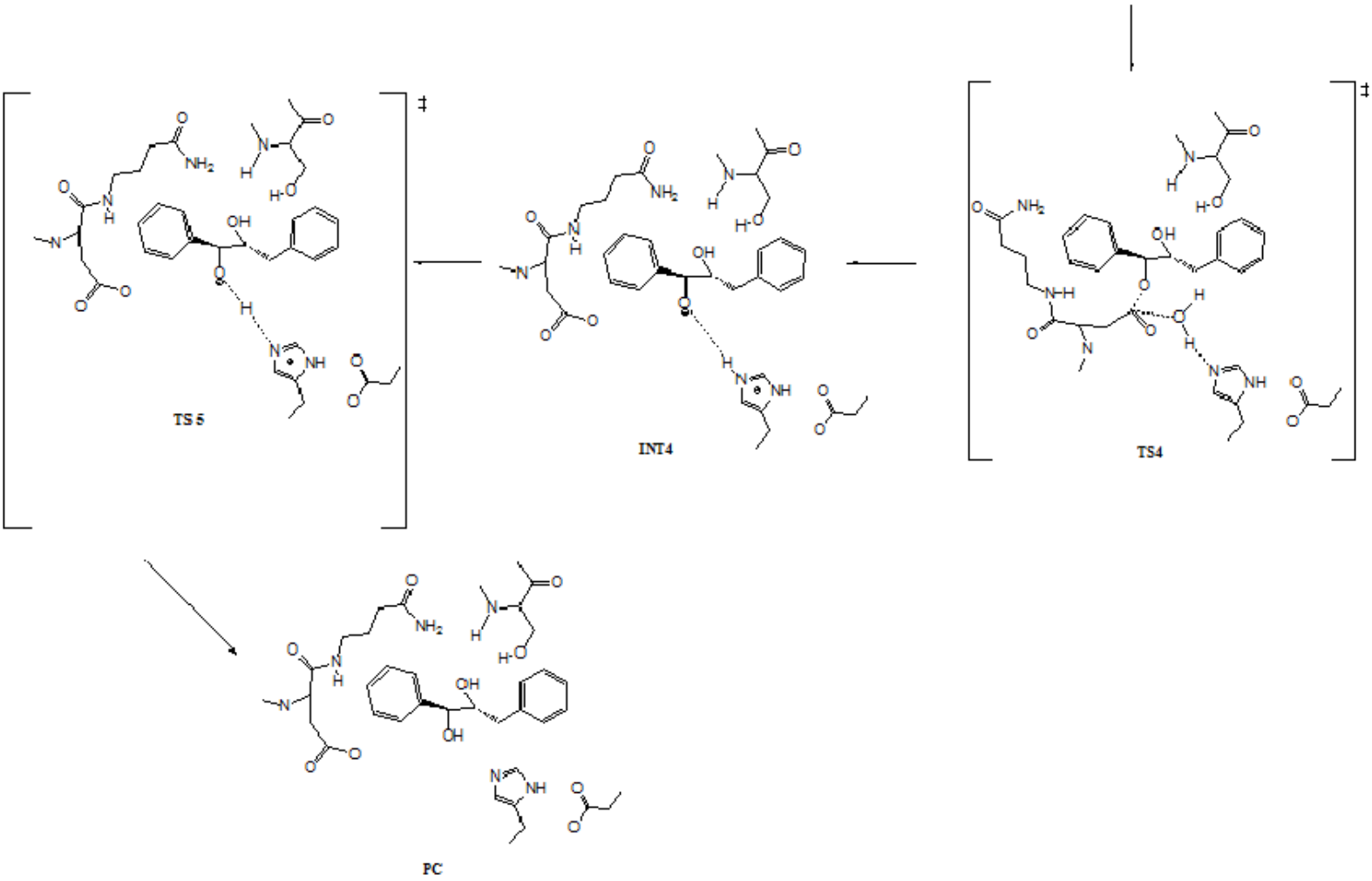


As we can see by simple inspection of Figure 19, the enzyme attacks reverently the Conformation 1 instead of the Conformation 2, and there is a energetic preference for Carbon 1 rather than for Carbon 2. Hence, the rest of the study of the Mechanism B will be presented only for C1 in Conformation 1, as is senseless to study the other Conformation as well as C2, as the attack will be preferentially on (1)-C1.

As has been previously mentioned, the first step was studied experimentally, but the rest of the mechanism is unclear. The major question is the origin of the hydrogen which binds to the oxygen of the epoxide after the formation of oxyanion intermediate. There are two possible origins for this hydrogen, which we will name B(2) and B(3) both depicted in the following scheme:

Scheme 8: Schematic representation of the two possible scenarios for Mechanism B.





-Ser105Asp Mutation. Study of the second step of Mechanism B, scenario B(2).

For the study of this scenario, the strategy that had been employed was the use of a 3D PES to localize TS3 and after that a 2D PES to arrive to products, going through TS4.

For the 3D PES, the following equations had been employed:

X-Axe: normal combination of distances:

$$r1 = d(C^{Asp} - O^{wat}) \quad (20)$$

Y-Axe: Antisymmetric combination of distances:

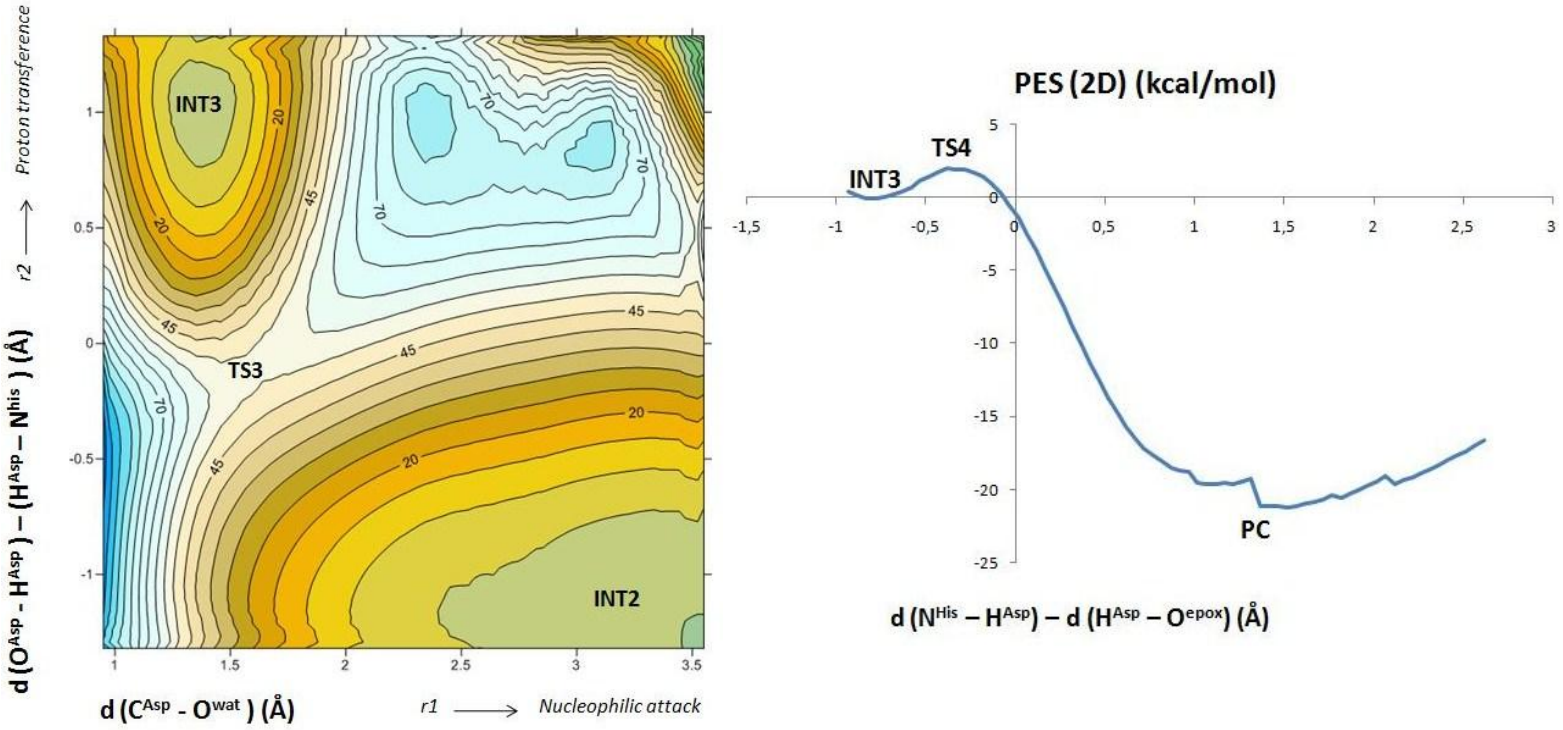
$$r2 = d(O^{Asp} - H^{Asp}) - d\,(H^{Asp} - N^{His}) \quad (21)$$

While for the 2D PES, the employed equation is the following antisymmetric combination of distances:

$$r1 = d\left(N^{His} - H^{Asp}\right) - d\,(H^{Asp} - O^{epox}) \quad (22)$$

The following figure shows the two employed potential energy surfaces for the study of this scenario:

Figure 20: 3D and 2D PES for the study of scenario B(2) for the nucleophilic attack to C1 of the t-DPPO (Conformation 1).



-Ser105Asp Mutation. Study of the second step of Mechanism B, scenario B(3).

For the study of this scenario, a different strategy had been undertaken. For the localization of TS3, a 2D PES had been employed. The next step of the scenario had been explored by means of a 3D PES, which allowed the localization of TS4. Finally, to arrive to products, a 2D PES had been used.

For the first 2D PES, the employed equation is the following antisymmetric combination of distances:

$$r1 = d\left(N^{His} - H^{Asp}\right) - d\left(H^{Asp} - O^{epox}\right) \quad (23)$$

For the second step of the scenario, the 3D PES, the following equations had been employed:

X-Axe: normal combination of distances:

$$r1 = d(C^{Asp} - O^{wat}) \quad (24)$$

Y-Axe: Antisymmetric combination of distances:

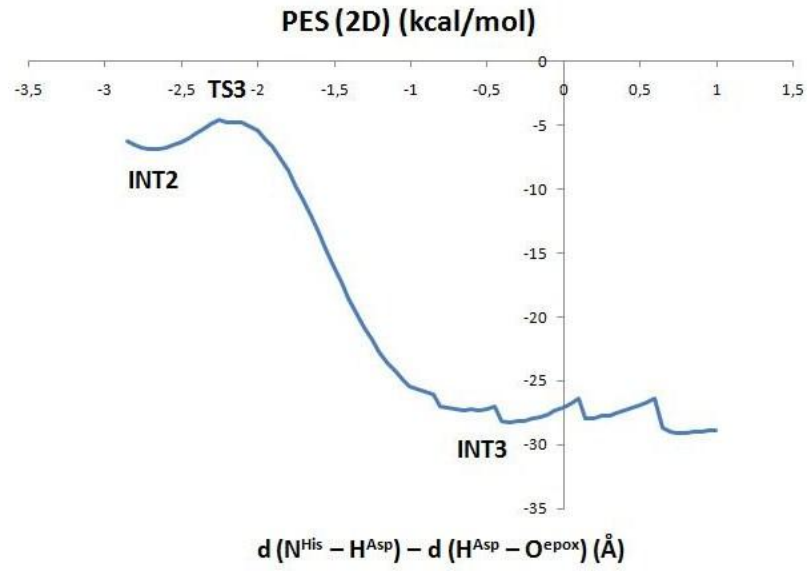
$$r2 = d(O^{wat} - H^{wat}) - d\left(H^{wat} - N^{His}\right) \quad (25)$$

Finally, for the location of TS5 and the end of the reaction, the equation of the 2D PES that has been used is:

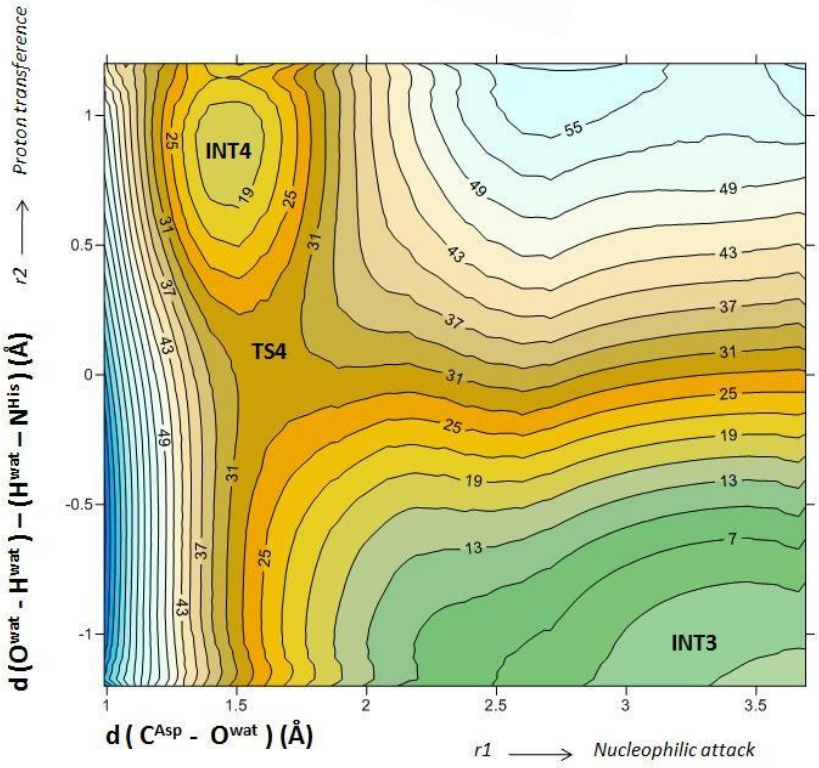
$$r1 = d\left(N^{His} - H^{wat}\right) - d\left(H^{wat} - O^{Asp}\right) \quad (26)$$

The following figure shows the employed potential energy surfaces for the study of this scenario:

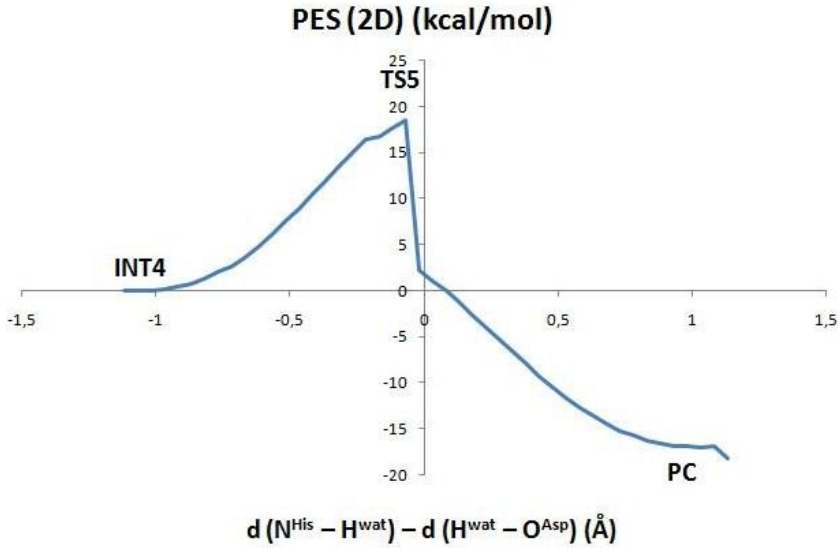
Figure 21: Potential energy surfaces employed for the study of scenario B(3) for the nucleophilic attack to C1 of the t-DPPO (Conformation 1). Figure 21(a) is the first step, the transfer of the proton from His224 to the O of the epoxide. Figure 21 (b) shows the nucleophilic attack of the water to the Asp105 and the transfer of one proton of the water to the His224. Finally, Figure 21 (c) shows the third and last step, the transfer of the proton of the water, taken by the His224, to the O of the Asp105, which had attacked the C of the epoxide in the first step of the mechanism, as described in Scheme 7.


PES (2D) (kcal/mol)
TS3
INT2
INT3
d (N^His – H^Asp) – d (H^Asp – O^epox) (Å)

(a)


INT4
TS4
INT3
d (O^wat - H^wat) – (H^wat – N^His) (Å)
Proton transference
r2
d (C^Asp - O^wat) (Å)
r1
Nucleophilic attack

(b)


PES (2D) (kcal/mol)
TS5
INT4
PC
d (N^His – H^wat) – d (H^wat – O^Asp) (Å)

(c)

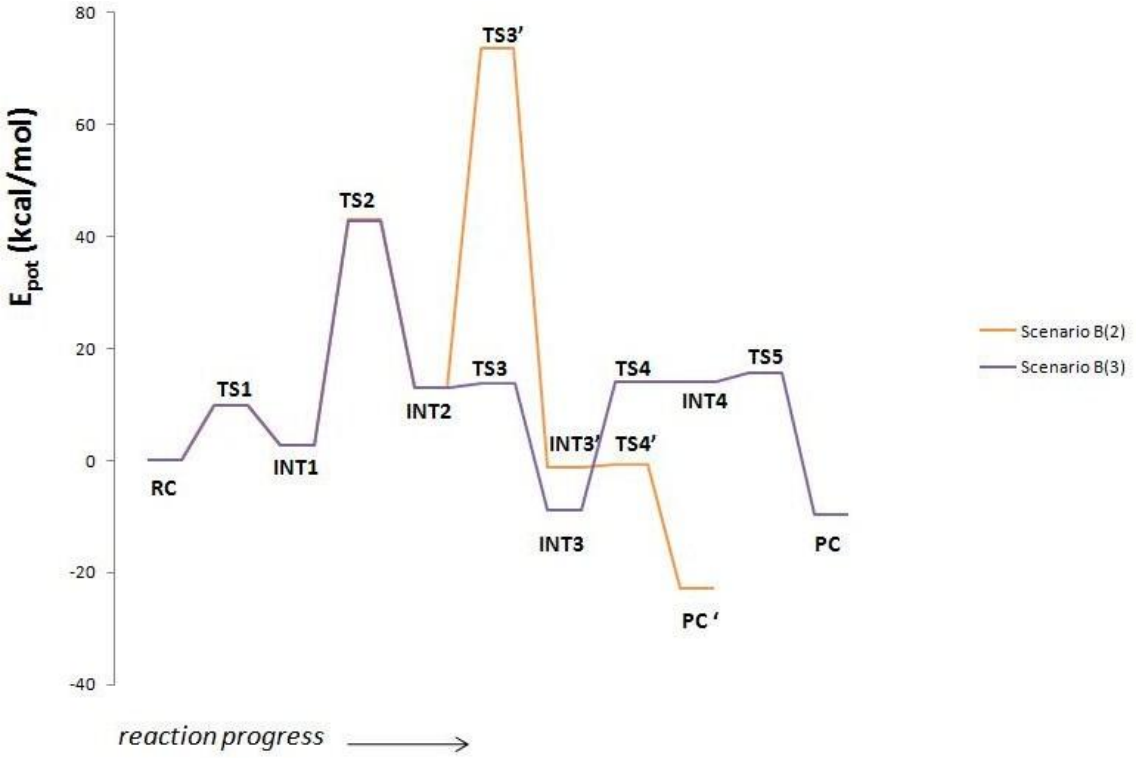
The following table contains a summary of the Potential energy values for each structure of both scenario B(2)and B(3):

Table 4: Potential Energy values (in kcal/mol) for the structures of scenario B(2) and scenario B(3) for conformation 1, attack on C1.

Structure	Scenario B(2)	Scenario B(3)
RC	0.0	0.0
TS1	9.8	9.8
INT1	2.6	2.6
TS2	42.9	42.9
INT2	13	13
TS3	73.5	13.7
INT3	-1.3	-9
TS4	-0.8	14
INT4	-	14
TS5	-	15.6
PC	-22.9	-9.7

Table 4 may be plotted out for a more visual approach to the energy barriers:

Figure 22: Potential energy surfaces employed for the study of scenario B(3) for the nucleophilic attack to C1 of the t-DPPO (Conformation 1).



4.3. - Study of Mechanism B on wild-type CAL-B.

As had been previously said in the introduction, it is necessary to perform a mutation on wild-type CAL-B to have Mechanism B available. This means that wild-type CAL-B cannot perform the nucleophilic attack by itself.

Then, Ser 105 is not a good nucleophile of the reaction. To prove this, a theoretical study have been carried out as well, only for C1 of the Conformation 1 of the epoxide, t-DPPO.

The first step of the mechanism has been done again, this time with wild-type CAL-B. The starting system that has been employed is the system shown in Figure 8. The mechanism that have been explored is Mechanism B(1), depicted in Scheme 7 and experimentally proved by Mulholland[63]. Hence, the strategy is exactly the same that had been previously used: first one 3D PES will be explored. After that, a 2D PES surface will be carried out, starting from Intermediate 1.

For the 3D PES, the following equations had been employed:

X-Axe: normal combination of distances:

$$r1 = d(O^{Ser} - C_i) \qquad (27)$$

Y-Axe: Antisymmetric combination of distances:

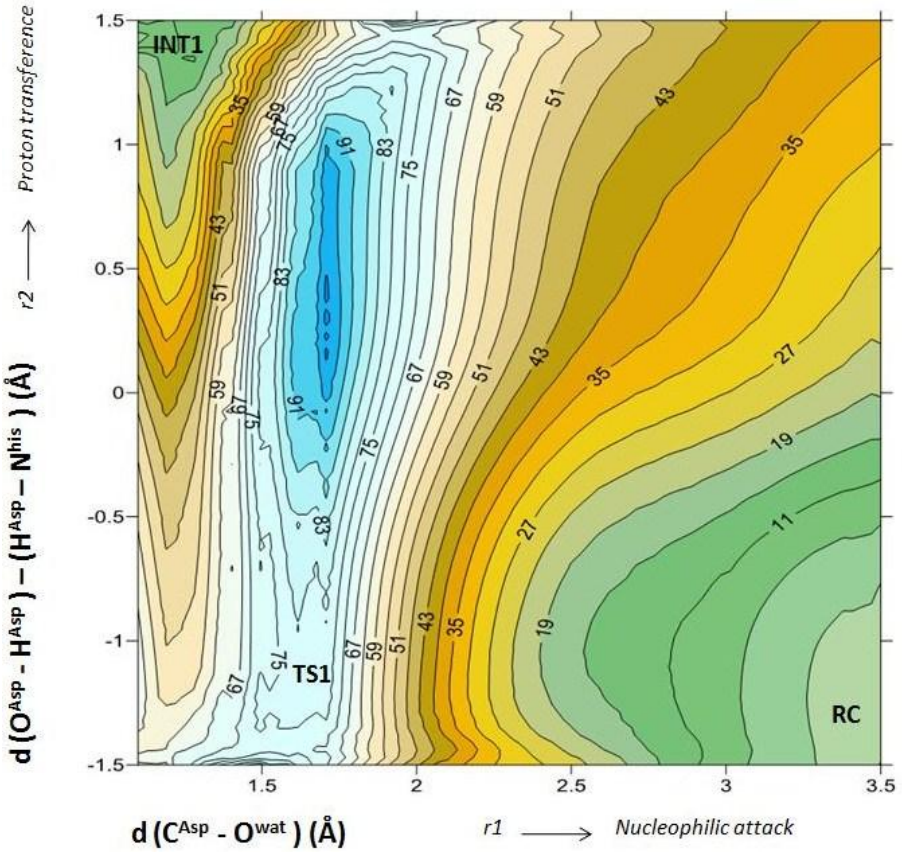
$$r2 = d(O^{Ser} - H^{Ser}) - d\,(H^{Ser} - N^{His}) \qquad (28)$$

Whilst for the 2D PES, the equation corresponds to the following antisymmetric combination of distances:

$$r1 = d(O^{epox} - C_i) - d\,(C_i - O^{Ser}) \qquad (29)$$

The following figure shows the 3D PES:

Figure 23: 3D Potential Energy Surface employed for the study of the nucleophilic attack of Ser 105 to C1 of the t-DPPO.



As it can be seen in Figure 23, the transition state of the reaction, labelled TS1, is in a very high energy area, circa 75 kcal/mol. This means that the reaction will never proceed by this path so the initial thoughts were correct and the mutation is required.

5. – Conclusions

At the beginning of this BDFG, was stated that the main goal was to elucidate the full mechanism of hydrolysis of DPPO. To do so, Mechanism A and Mechanism B had been explored in all the possible scenarios: two conformations and the attack to two C of the epoxide. Then, the first thing to find out is which mechanism is the chosen one. The Mechanism A (see Scheme 6) corresponds to the overture of the epoxide by means of a nucleophilic attack of a water molecule whilst Mechanism B (see Schemes 6 to 8) corresponds to the nucleophilic attack of a residue of the enzyme to a carbon of the epoxide.

If Figure 14 and Figure 19 are compared, it can be seen that Mechanism A has a larger energy value for TS1 than Mechanism B. This means that Mechanism B is preferred over Mechanism A, as the energy is too high and the reaction will not proceed in such way.

So Mechanism B is the mechanism of the reaction. Now, what has to be found out is which conformation and carbon for the nucleophilic attack is preferred. This can be done by inspection of Figure 19. As it can be seen there, the rate-limiting step[64] is the conversion from INT1 to INT2, going through TS2. Figure 19 shows that the nucleophilic attack to C1 of t-DPPO *i.e.* Conformation 1 is energetically preferred, as it has the lowest energetic barrier.

Henceforth, we know that the reaction mechanism will be the one in which the nucleophilic attack is performed by a residue of the enzyme. Moreover, it has been found out that within this mechanism, the reaction will proceed by the nucleophilic attack to the C1 of the t-DPPO 1 also known as Conformation 1 of the epoxide.

The only thing that remains unclear is which scenario is energetically favourable and so, the one by means of which the reaction will proceed. There are two possible scenarios which essentially differ in the moment in which the proton which binds to the oxygen of the epoxide will be transferred. One scenario, called Mechanism B(2) assumes that the proton will be transferred after the nucleophilic attack of the Ser105Asp to the carbon of the epoxide. The other scenario, labelled B(3) suggests that the proton of the Ser105Asp is transferred in before the nucleophilic attack.

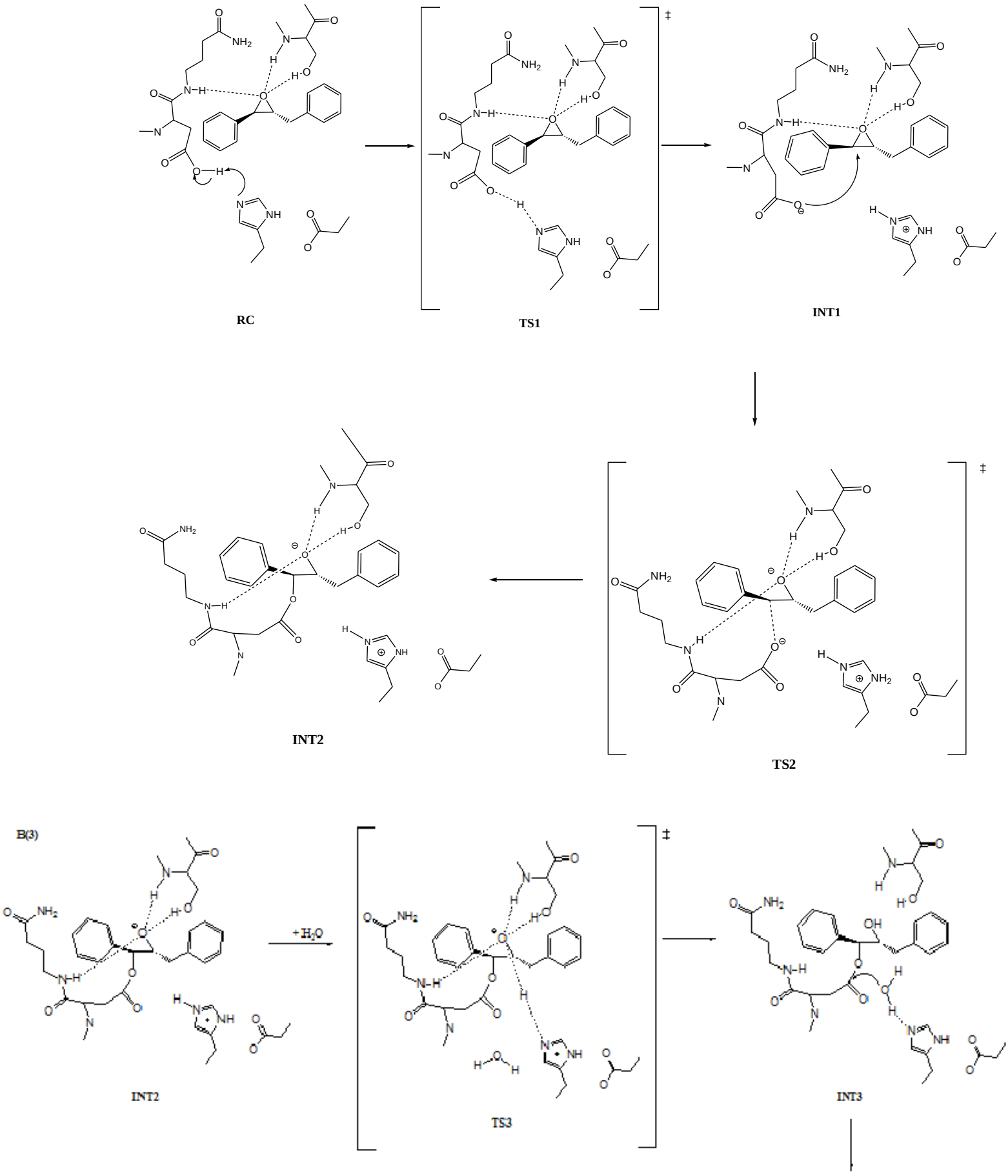
To know whether one or another is the correct, the inspection of Figure 22 is enough.

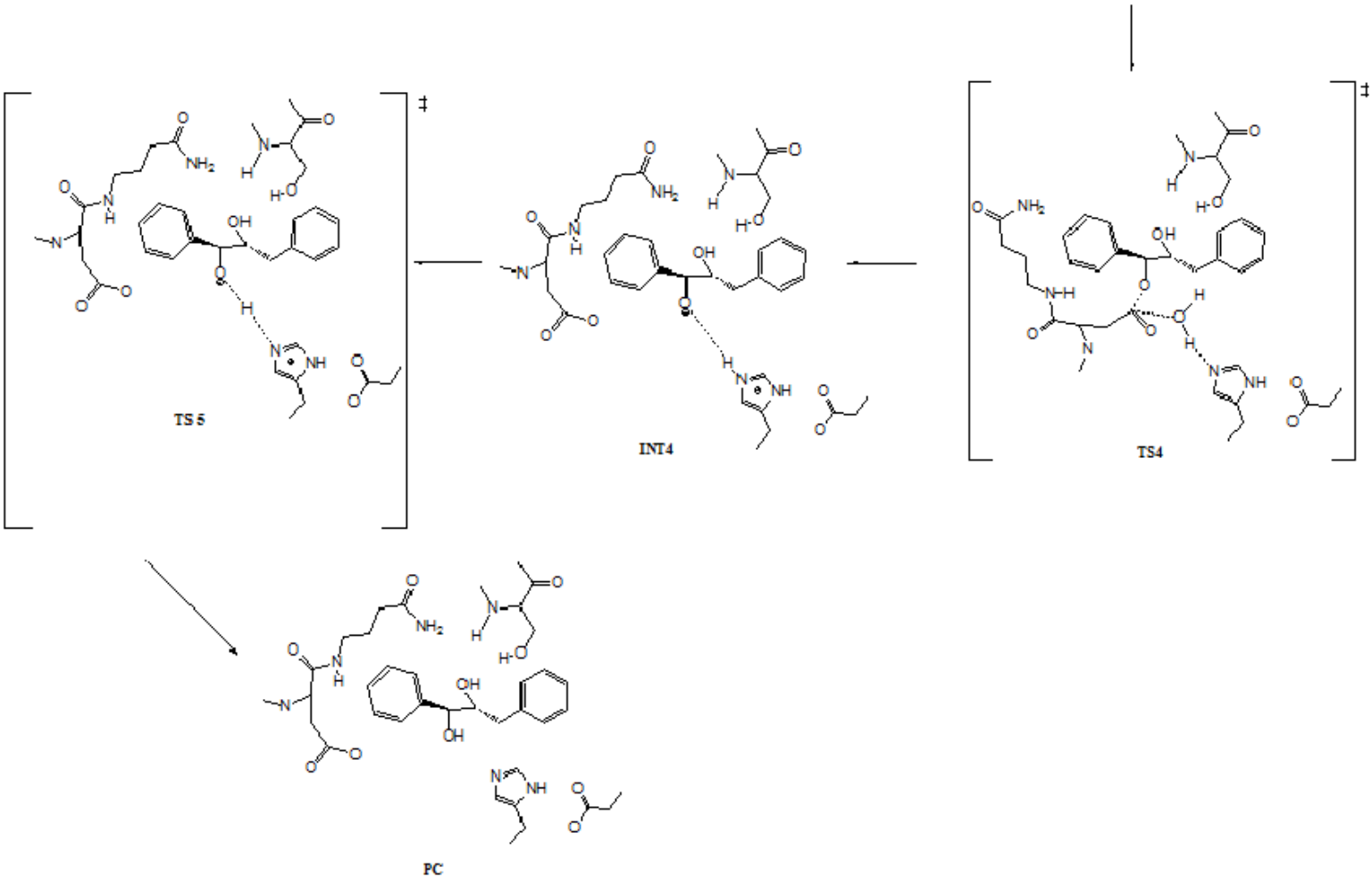
As it can be easily seen there, scenario B(2), where the nucleophilic attack of the water is first and then the transfer of the pro has much higher barriers for TS3 than scenario B(3) has.

Attending to the previous reasoning, it can be stated that the reaction will proceed by means of Mechanism B, scenario B(3). Then, the question of the origin of the proton that was not elucidated by Mulholand and co-workers is now clear: the proton which binds to the oxygen of

the epoxide after the formation of oxyanion intermediate comes from the mutated Asp105, and the mechanism of the reaction will be the one depicted in the next scheme:

Scheme 9: Multistep mechanism of hydrolysis of t-DPPO catalyzed by CAL-B.





Then, the mutation proposal that had been done worked very well, so the side objective of proposing mutations to increase the secondary catalytic efficiency had been accomplished as well.

Unfortunately, one of the side aims cannot be answered. At the level of theory of this BDFG, QM AM1, the enantioselectivity of the enzyme cannot be observed, as there is no clear distinction within the different barriers for both carbons.

This Bachelor's Degree Final Project has been included in a manuscript that has been submitted for publication in the scientific journal Molecules for its publication at the end of July.

6. – Further work

One possible choice to elucidate whether the enzyme shows enantioselectivity is to carry out studies with higher theory level. For instance, instead of using QM methods, QM/MM methods can be used to the study of the reaction, employing the semi-empirical AM1 Hamiltonian.

Another thing that can be done to a better prediction is to carry out dynamics analysis, in order to calculate free energies instead of studying only potential energies.

All these things will be done in a brand new paper in which we are currently working in which publication date may be by the end of August of this year.

7. – Acknowledgements

First of all, I would like to thank my tutor, Prof. Vicente Moliner for giving me the chance of doing my Bachelor's Degree Final Project in his research group. It was wonderful to work with such incredible people. I want to thank Dr. Katarzyna Świderek for her multiple advises in the making of this BDFG as well as many group meetings which helped me to know how scientists work: as a block instead of a brick.

During these 5 months of work the BDFG I have been all the time kindly accompanied and advised by the future Dr. Isabel Bordes Pastor. I would like to thank her for all the things that she has done: from teaching me the basics of UNIX language to staying until 10 p.m. with me in the laboratory localizing a nasty transition state.

I would like to express my heartiest gratitude to Dr. Javier Ruiz Pernía and Sergio Martí Forés for their many explanations in both Physical Chemistry fundamental and Computational methods, tricks and results analysis. Thanks also to Erica, Kemel, Nacho and the other members of the group that, somehow, helped me in my time as Computational Chemist.

During this 5 years I have attended many lectures and been taught by even more teachers. These very brief lines are dedicated to them, to their effort in not only teaching me but getting to love what I was doing. You are doing very well, please, keep going.

This BDFG means the end of a very nice period of my life, and so, I want to thank my colleges and classmates, the ones who remain and the ones who left, who have been by my side, independently of my situation. Thank to them, this 5 years had been better.

To finish this section, my last but not least acknowledgement to my family, to the ones that are with me as well as to the ones that, unfortunately, had to leave me before this moment, to all of you, thanks.

Within this I would like to give special thanks to my parents, Pepe and Maria Teresa as well as to my younger brother Miguel, for their constant support at every single point.

Being the previous words already spoken, I would like to finish this BDFG as well as my time as Undergraduated Student with some word of one of my favourite books: Faust by Johann Wolfgang von Goethe with the hope that they can also inspire other people as well as they did with me:

,,Vi veri ueniversum vivus vici''

8. – References

(1)Cox, D. L. N. a M. M.*Lehninger Principles of Biochemistry*, 5th ed. New York, (2008).

(2) Kohen, A. *Science Reviews* **(2003)**, 28.

(3) Gotor, V., Alfonso, I. and Garcîa -Urdiales, E. (**2008**). *Asymmetric organic synthesis with enzymes*, Wiley-VCH, Weinheim, **2008**.

(4) a) G. Carrea, S. Riva, *Organic Synthesis with Enzymes in Non-Aqueous Medium*, Wiley-VCH, Weinheim, 2008; b) P. Lozano, *Green. Chem.* 2010, 12, 555.569; c) M. J. Hernáiz, A. R. Alcántara, J. I. García, J. V. Sinsisterra, *Chem. Eur. J.* 2010, 16, 9422.9437.

(5) a) S. L. Y. Tang, R. L. Smith, M. Poliakoff, *Green. Chem.,* 2005, 7, 761-762; b) P. Anastas, N. Eghbali, *Chem. Soc. Rev.,* 2010, 39, 301-312.

(6) a) W. D. Fessner, T. Anthosen, *Modern Biocatalysis. Stereoselective and Environmentally Friendly Reactions,* Willey-VCH, Weinheim, 2009; b) T. Hudlicky, J. W. Reed, *Chem. Soc. Rev.*2009, 38, 3117-3132.

(8) a) J. M. Woodley, *Trends Biotechnol.,* 2008, 26, 321-327; b) R. N. Patel, *Coord. Chem. Rev.,* 2008, 252, 659-701; c) J. Tao, J.-H Xu, *Curr. Opin. Chem. Biol.,* 2009, 13, 43-50.

(9) Elliott, W. and Elliott, D. (1997). *Biochemistry and molecular biology*. Oxford: Oxford University Press.

(10) Price, N. and Stevens, L. (1982). *Fundamentals of enzymology*. Oxford: Oxford University Press.

(11) Voet, D. and Voet, J. (1995). *Biochemistry*. New York: J. Wiley & Sons.

(12) Price, N. and Price, N. (2001). *Principles and problems in physical chemistry for biochemists*. Oxford: Oxford University Press.

(13) Frey, P. and Hegeman, A. (2007). *Enzymatic reaction mechanisms*. New York: Oxford University Press.

(14) Nelson, D., Nelson, D., Lehninger, A. and Cox, M. (2008). *Lehninger principles of biochemistry*. New York: W.H. Freeman.

(15) Fischer, E *Ver. Dt. Chem. Ges.,* **1984**, 27.

(16) Figure taken from the lecture notes of Dr. Raquel Castillo Solsona, given in the unit *QU0921: Bioquímica* of 2013.

(17) Michaelis, L., Menten, M. L., *Biochemische Zeitschriff*, **1913**, 49, 333.

(18) Haldane, J. B. S., *Enzymes*, **1930**, Longmans, Green and Co.

(19) Pauling, L. *Chem. Eng. News,* **1946**, 24, 1375.1377.

(20) Martí, S., Roca, M., Andres, J., Moliner, V., Silla, E., Tuñón. I., Bertran, J., *Chemical Society Rewiews*, **2004**, 33, 98.

(21)Svedendahl, M.; Jovanovic, B.; Fransson, L.; Berglund, P., *ChemCatChem* **2009**, 1, 252−258.

(22)Katarzyna Świderek, Sergio Martí, and Vicent Moliner., *ACS Catal*. **2014**, 4, 426−434.

(23) a) D. L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, J. Schrag, J. L. Sussman, K. H. G. Verschueren, A. Goldman, *Protein Eng.,* **1992**, 5, 197-211; b) M. Holmquist, *Curr. Protein Pept. Sci*., **2000**, 1, 209-235; c) Świderek, K.; Pabis, A.; Moliner, V.; *Org. Biomol. Chem*., **2012**, 10, 5598−5605.

(24) Busto, E., Gotor-Fernández, V., Gotor, V., *Chem. Soc. Rev.*, **2010**, 39, 4504-4523.

(25) a) Kirk, O., Christensen, M. W., *Org. Process Res..* **2002**, 6, 446– 451. b) Anderson, E. M., Larsson, K. M.; Kirk, O., *Biocatal. Biotransform.* **1998**, 16, 181–204. c) Kirk, O., Björkling, F., Godtfredsen, S. E., Larsen, T. O., *Biocatalysis* **1992**, 6, 127–134.

(26) a) Gross, R. A.; Kalra, B. *Science* **2002**, 297, 803–806. b)Gross, R. A.; Kumar, A.; Kalra, B. *Chem. ReV*. **2001**, 101, 2097– 2124. c)Kumar, A.; Gross, R. A. *J. Am. Chem. Soc.,* **2000**, 122, 11767–11770. d) Mahapatro, A.; Kumar, A.; Kalra, B.; Gross, R. A. *Macromolecules* **2004**, 37, 35–40. e) Chen, B., Hu, J., Miller, E., Xie, W., Cai, M. and Gross, R.; **2008,**. *Biomacromolecules*, 9(2), 463-471.

(27) a) O'Brien, P.J., Herschlag, D., *Chem. Biol.,* **1999**, 6, R91-R105; b) Aharoni, A., Gaidukov, L., Khersonsky, O., McQ Gould, S., Roodveldt, C., Tawfik D. S., *Nat. Genet. ,* **2005**, 37, 73-76; c) Hult, K., Berglund, P., *Trends Biotech.,* **2007**, 25, 231-238.

(28) Svedendahl, M., Carlqvist, P., Branneby, C., Allner, O., Frise, A., Hult, K., Berglund, P., Brinck, T., *ChemBioChem*, **2008**, 9, 2443–2451.

(29) Morisseau, C., Hammock, B., **2005**, *ChemInform*, 36(39).

(30) a) Archelas A, Furstoss R. **2001**. *Curr.Opin. Chem. Biol.* 5:112–19; b) de Vries EJ, Janssen DB. **2003**. *Curr. Opin. Biotechnol.* 14:414–20.

(31) Wixtrom RN, Hammock BD. **1985**. Membrane-bound and soluble-fraction epoxide hydrolases: methodological aspects. In Biochemical Pharmacology and Toxicology: Methodological Aspects of Drug Metabolizing Enzymes, edDZakim, DA Vessey, vol. 1, pp. 1–93. New York: Wiley.

(32) a) Yu, Z., Xu, F., Huse, L., Morisseau, C., Draper, A., Newman, J., Parker, C., Graham, L., Engler, M., Hammock, B., Zeldin, D. and Kroetz, D., **2000**. *Circulation Research*, 87(11), 992-998; b) Zhao X, Yamamoto T, Newman JW, Kim IH,Watanabe T, et al. **2004**, *J. Am. Soc. Nephrol.* 15:1244–53; c) Imig, J., Zhao, X., Capdevila, J.,Morisseau, C. and Hammock, B. **2002**. *Hypertension*, 39(2),690-694.

(33) Holmquist, M., *Curr Protein Pept Sci*. **2000**, 1(2):209-35.

(34)Beetham, J., Grant, D., Arand, M., Garbarino, J., Kiyosue, T., Pinot, F., Oesch, F., Belknap, W., Shinozaki, K. and Hammock, B., **1995**. *DNA and Cell Biology*, 14(1), 61-71.

(35) Yamada, T., Morisseau, C., Maxwell, J.E., Argiriadi, M. A., Christianson, D. W., et al. **2000**., *J. Biol. Chem.* 275:23082–88.

(36) Jerina, D. M., **1994**. *Lloydia* 37:2012-18.

(37)Oliw, E.H., **1994**. *Prog. Lipid Res.* 33:329-54.

(38) a) Szeliga J, Dipple A. **1998,** *Chem. Res. Toxol.* 11, 1-11; b) Zheng J, Myung C, Jones AD, Hammock BD. **1997**, *Chem. Res. Toxicol.* 10, 1008-1014.

(39) Web page with a historical review of the Computational Chemistry: http://www.netsci.org/Science/Compchem/feature17a.html

(40) Hameka, H. F.,**2004**. Quantum mechanics. Hoboken, N.J.: Wiley-Interscience.

(41) P.A. Atkins and J. de Paula, *Atkins' Physical Chemistry, 7th Edition, Oxford University Press 2002.*

(42) Steiner, E. 2008. The chemistry maths book. 2nd ed. Oxford: Oxford University Press.

(43) Born, M.; Oppenheimer, R. Ann. Phys. (Leipzig) **1927**, 84, 457-484.

(44) Anslyn, E. V. and Dougherty, D. A. **2006**. Modern physical organic chemistry. Sausalito, CA: University Science.

(45)Warshel, A. and Levitt, M. **1976**. *Journal of Molecular Biology*, 103 (2), 227-249.

(46) Field, M. J., *A Practical Introduction to the Simulation of Molecular Systems*; Cambridge University Press: Cambridge, UK, **2007.**

(47) a) Jorgensen W.L. and Tirado-Rives J., **1988**,*J. Am. Chem. Soc.,* 110, 1657. b) Jorgensen W.L., Maxwell D.S. and Tirado-Rives J., **1996**, *J. Am. Chem. Soc.*, 118, 11225. c) Kaminski G.A. and Friesner R.A., Tirado-Rives J. and Jorgensen W.L., **2001**, *J. Phys. Chem. B*, 105, 6474.

(48) a) Weiner ,S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. Jr., Weiner P., *J. Am. Chem. Soc.,* 106, 765, **1984**. b) Weiner, S.J., Kollman, P.A., Nguyen, D.T., Case, D.A., *J. Comput. Chem.,* 7, 230, **1986**. c) Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M. Jr., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., Kollman, P.A., *J. Am. Chem. Soc.,*117, 5179, **1995**.

(49) a) McKerell, A.D., Banavali, N.K.,**2000**. *J. Comp. Chem.*, 21, 105. b) Patel, S., Brooks, C.L., **2004.** *J. Comput. Chem.,* 25*,* 1.

(50) Jensen, F. 1999. Introduction to computational chemistry. Chichester: Wiley.

(51) Warshel A., Levitt, M. *Journal of Molecular Biology,* **1976**, 103, 227.

(52) Figure adapted from: A. R. Leach, Molecular Modelling, 2nd Ed., Prentice Hall, 2001, p. 284.

(53) Levine, I., **2000**. *Quantum chemistry*. Upper Saddle River, N.J.: Prentice Hall.

(54) J. Uppenberg, M. T. Hansen, S. Patkar, T. A. Jones, *Structure*, **1994**, *2*, 293-308.

(55) a) Sondergaard, Chresten R., Mats HM Olsson, Michal Rostkowski, and Jan H. Jensen., *Journal of Chemical Theory and Computation 7*, no. 7, **2011**: 2284-2295. b) Olsson, Mats HM, Chresten R. Sondergaard, Michal Rostkowski, and Jan H. Jensen. *Journal of Chemical Theory and Computation 7*, no. 2, **2011**: 525-537.

(56) GaussView, Version 5, Dennington, R.; Keith, T.; Millam, J. *Semichem Inc.*, Shawnee Mission KS, **2009**.

(57) M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, J. Am. Chem. Soc., **1985**, 107, 3902.

(58) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey R. W., Klein M. L., *J. Chem. Phys.* **1983**, 79, 926.

(59) M. J. Field, M. Albe, C. Bret, F. Proust-De Martin, A. Thomas, *J. Comp. Chem.* **2000**, 21, 1088-1100.

(60) DassaultSystèmes BIOVIA, Discovery Studio Modeling Environment, Release 4.5, San Diego: DassaultSystèmes, **2015**.

(61) Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., Rosenberg, J. M., *J. Comp. Chem.*, **1992**, 13, 1011.

(62) Torrie, G. M., Valleau, J. P., *J.Comp. Phys.***1977**, 23, 187.

(63) Richard Lonsdale, Simon Hoyle, Daniel T. Grey, Lars Ridder, and Adrian J. Mulholland., *Biochemistry* **2012**, 51, 1774−1786.

(64) Murdoch, J. (**1981**), *J. Chem. Educ.*, 58(1), 32.

(65) All chemical compounds and mechanisms have been drawn by means of ChemDraw 9.0 Ultra: Cousins, K., **2005**. *J. Am. Chem. Soc*., 127(11), 4115-4116.