Universitat de Girona

Facultat de Ciències

Memòria del Treball Final de Grau

Títol del treball:

The secrets of biocatalysis: adaptation of epoxide hydrolase enzyme from *Bacillus megaterium* for drug synthesis

Estudiant: **JUDIT TORRENT PALAHÍ**

Grau en **QUÍMICA**

Correu electrònic: juditp93@gmail.com

Tutor: Dr. Sílvia Osuna Oliveras

Cotutor*:

Empresa/Institució: Institut de Química Computacional i Catàlisi, UdG

Vistiplau tutor (i cotutor*):

Nom del tutor: Dr. Sílvia Osuna Oliveras
Nom del cotutor*:
Empresa/Institució: IQCC, UdG
Correu(s) electrònic(s): silvia.osuna@udg.edu

*si hi ha cotutor assignat

Data de dipòsit de la memòria a secretaria de coordinació: 19th JULY, 2018

Universitat de Girona

THE SECRETS OF BIOCATALYSIS:

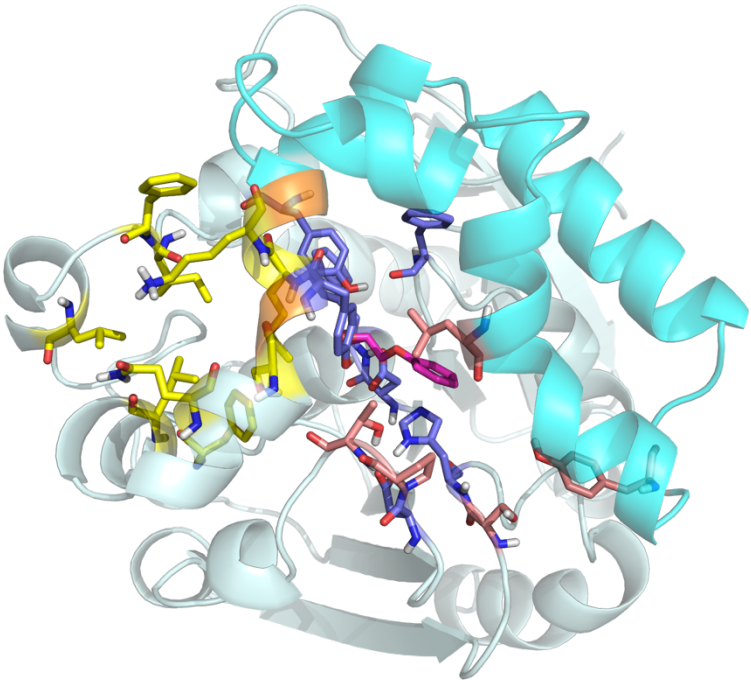
ADAPTATION OF EPOXIDE HYDROLASE ENZYME FROM *BACILLUS MEGATERIUM* FOR DRUG SYNTHESIS

Chemistry Research Project

Supervised by Dr. Sílvia Osuna (IQCC, UdG)

JUDIT TORRENT PALAHÍ

19th July, 2018



iQCC
Institut de Química Computacional i Catàlisi

CompBioLab

RESUM

Des de l'inici de la vida es poden trobar a la natura unes biomolècules que realitzen moltes de les reaccions químiques necessàries per a la vida. Aquestes biomolècules són conegudes com enzims, i de fet son els catalitzadors més eficients, específics i selectius que hi ha. Això ha portat a la biocatàlisi, que es basa en l'ús d'enzims per catalitzar la transformació de compostos orgànics. Tot i els grans avantatges, hi ha moltes reaccions d'interès industrial que no es poden dur a terme eficientment utilitzant directament els enzims que trobem a la natura. Això ha conduït a l'enginyeria enzimàtica, que es basa en modificar i utilitzar els enzims per a nous propòsits pels quals no s'havien dissenyat originalment.

Tal i com s'ha esmentat anteriorment, els enzims que es troben a la natura no solen complir amb els requisits de les indústries farmacèutiques i de la química fina. En alguns casos pot ser degut a la seva baixa regioselectivitat i l'enantioselectivitat respecte als substrats no naturals d'interès industrial. En d'altres, degut al fet que l'enzim natural no accepta un ampli rang de substrats, la qual cosa, per exemple, no permetria la bioresolució de substrats més voluminosos. Aquest podria ser un dels principals inconvenients relacionats amb l'enzim epòxid hidrolasa (EH) del *Bacillus megaterium* (*Bm*EH), l'enzim d'interès en aquest projecte. L'enzim *Bm*EH mostra activitat cap a substrats voluminosos com ara, el naftil glicidil èter (NGE) (precursor del beta propranolol) però tot i això l'eficiència és pitjor en comparació a les observades pel substrat natural fenil glicidil èter (PGE).

En general, les EHs són un tipus d'enzim que formen part de la família α,β–hidrolasa capaç de realitzar hidròlisis d'epòxids de forma selectiva. En particular, la *Bm*EH catalitza preferentment l'obertura del grup epòxid de l'enantiòmer *(R)* del PGE, deixant l'enantiòmer *(S)* idealment sense reaccionar, la qual cosa és interessant perquè l'enantiòmer *(S)* és precisament l'intermedi valuós del beta-bloquejant alprenolol. Tot i que actualment es coneix el mecanisme general de les EHs, encara no es coneixen detalls de processos que ocorren abans (unió del substrat) i/o després (alliberament del producte). Per això, l'objectiu d'aquest estudi es comprendre la dinàmica conformacional de l'enzim natural *Bm*EH en presència de cada enantiòmer del PGE (sistema *Bm*EH – *(R/S)*-PGE) mitjançant simulacions de dinàmica molecular (MD).

RESUMEN

Desde el inicio de la vida se pueden encontrar en la naturaleza unas biomoléculas que realizan muchas de las reacciones químicas necesarias para la vida. Estas biomoléculas, conocidas como enzimas, son de hecho, los catalizadores más eficientes, específicos y selectivos que hay. Esto ha llevado a la biocatálisis, que se basa en el uso de enzimas para catalizar la transformación de compuestos orgánicos. A pesar de las grandes ventajas, hay muchas reacciones de interés industrial que no se pueden llevar a cabo eficientemente utilizando las enzimas que encontramos directamente en la naturaleza. Esto ha conducido a la ingeniería enzimática, que se basa en modificar y aplicar enzimas para nuevos propósitos para los que no se habían diseñado originalmente.

Tal y como se ha mencionado anteriormente, las enzimas que se encuentran en la naturaleza no suelen cumplir con los requisitos de las industrias farmacéuticas y de las químicas finas. En algunos casos puede ser debido a su baja regioselectividad y la enantioselectividad respecto a los sustratos no naturales de interés industrial. En otros, debido a que la enzima natural no acepta un amplio rango de sustratos, lo que, por ejemplo, no permitiría la bioresolución de sustratos voluminosos. Este podría ser uno de los principales inconvenientes relacionados con el epóxido hidrolasa del Bacillus megaterium (*Bm*EH), la enzima de interés en este proyecto. La enzima *Bm*EH muestra actividad hacia sustratos voluminosos tales como, el naftil glicidil éter (NGE) (precursor del beta propranolol) pero aún así la eficiencia es peor en comparación a las observadas por el sustrato natural fenil glicidil éter (PGE).

En general, las EH son un tipo de enzimas que forman parte de la familia α, β–hidrolasa capaz de realizar hidrólisis de epóxidos de forma selectiva. En particular, la *Bm*EH cataliza preferentemente la apertura del grupo epoxi del enantiómero *(R)* del (PGE), dejando el enantiómero *(S)* sin reaccionar, lo cual es interesante porque el enantiómero *(S)* es precisamente un intermedio valioso del alprenolol. Aunque actualmente se conoce el mecanismo general de las EHs, todavía no se conocen detalles de procesos que ocurren antes (unión del sustrato) y/o después (liberación del producto). Por ello, el objetivo de este estudio es comprender la dinámica conformacional de la enzima natural *Bm*EH en presencia de cada enantiómero del PGE (sistema *Bm*EH – *(R/S)*-PGE) mediante simulaciones de dinámica molecular (MD).

ABSTRACT

Since the beginning of life, it can be found in Nature biomolecules that make all the chemical reactions necessary for life. These biomolecules known as enzymes are indeed the most efficient, specific and selective catalyst known. This has led to biocatalysis, the use of enzymes to catalyse the transformation of organic compounds. Despite the great advantages, there are lots of reactions of industrial interest that cannot be carried out with high efficiencies by natural enzymes. This has led to enzyme engineering, whose job is to modify and apply laboratory evolved enzymes for new purposes for which they were not originally designed.

As mentioned above, enzymes found in Nature usually do not reach the requirements of pharmaceutical and fine chemical industries. In some cases, it may be due to its low regioselectivity and enantioselectivity towards non-natural substrates of industrial interests. In others, it may be caused by the narrow substrate scope of the WT enzyme, which, for example, does not allow the bioresolution of bulky substrates. This could be one of the main drawbacks related to epoxide hydrolases (EHs) enzyme of *Bacillus megaterium (Bm*EH), the enzyme of interest in this research work. *Bm*EH enzyme shows activity towards bulky substrates, such as naphthyl glycidyl ether (NGE) (valuable precursor of propranolol beta blocker drug), but nevertheless its efficiency is worse compared to those observed by the natural phenyl glycidyl ether (PGE) substrate.

In general, EHs are a type of enzymes belonging to the α,β-hydrolase family that are able of performing selective asymmetric epoxide hydrations. In particular, *Bm*EH catalyses the opening of the epoxide ring of the *(R)*-enantiomer of the PGE, leaving the *(S)* enantiomer ideally unreacted, which is interesting because the *(S)* enantiomer is a valuable intermediate beta-blocker of the alprenolol drug. Although the general mechanism of EHs is currently known, details of processes that occur before (substrate binding) and/or after (product release) are not yet known. For this reason, the aim of this study is to understand the conformational dynamics of the *Bm*EH natural enzyme in the presence of each enantiomer of PGE (*Bm*EH – *(R/S)*-PGE system) through Molecular Dynamic (MD) simulations.

I would like to express my deep gratitude to Dr. Sílvia Osuna and Ms. Eila Serrano for their patient guidance, enthusiastic encouragement, useful critiques and assistance in keeping my progress on schedule of this research work. I would also like to thank Mr. Guillem Casadevall and to all the researchers in CompBioLab (IQCC).

Finally, I wish to thank my parents, Lluís and M. Carme, my sister Sílvia, my uncle Jordi, my cousin Gerard and Elisabet, family also, my friends Diana and Anna for their support and encouragement throughout my work and studies.

July 19th, 2018

LIST OF ABBREVIATIONS

Abbreviation	Description
Å	Angström
AMBER	Assisted model building with energy refinement
atm	atmosphere
*Bm*EH	*Bacillus megaterium* epoxide hydrolase
C / CA	Carbon atom / alpha Carbon atom
E-S	Enzyme-substrate
EH	Epoxide hydrolase
ff99SB	Amber force field 99 Stony Brook modification
K	Kelvin
kcal	Kilo calorie
MD	Molecular Dynamic
MM	Molecular Mechanics
N	Nitrogen atom
NGE	Naphthyl glycidyl ether
ns	Nanosecond
OD1 Asp97	Side chain carbonyl oxygen atom of aspartic residue number 97 (*Bm*EH PDB entry: 4nzz numbering)
OH Tyr144, 203	Hydroxyl group of tyrosine residues number 144 and 203 (*Bm*EH PDB entry: 4nzz numbering)
PGE	Phenyl glycidyl ether
QM	Quantum mechanics
RMSd	Root mean square deviation
RMSf	Root mean square fluctuation
WT	Wild-type

CONTENTS

1. INTRODUCTION

1.1 Overview

Enzymes are large biomolecules found in Nature that are responsible of accelerating many chemical reactions that are necessary for life. Among all of their characteristics, some of them are of high relevance: they are highly efficient and specific, which means that they accelerate the rate of the reaction many orders of magnitude of one kind of compound by decreasing the activation barriers in a more proficient way. They operate under mild conditions and in water as solvent, are biodegradable, non-toxic, and decrease the total steps of chemical processes.[1] And this is what makes the application of enzymes in industry really interesting.

In early 20th century, pharmaceutical industry started to change as scientists began to study and compile information of existing processes in order to rationally understand enzymes much better. And so, this knowledge allowed to develop new mechanisms to synthesize organic compounds with high yields and high purities.[2] And soon the application of enzyme in industries in synthetic chemistry (known as biocatalysis), become very appealing. One of the first applications of biocatalysis in industry is the production of acetic acid from ethanol with an immobilised *Acetobacter* strain.[3]

Despite the advantages of biocatalysis, there are some key things that are still not solved. For instance, it is still a challenge to stabilise the enzyme at the desired pHs, temperatures and solvents employed in industries. And so, the use of enzymes as biocatalysts in pharmaceutical industry is sometimes affected by the low activity and selectivity (enantioselectivity and regioselectivity) of the wild-type (WT) enzyme towards non-natural substrates, or because enzymes suffer product inhibition, they have a narrow substrate scope,[4] and/or a lack of stability in organic solvents.

However, with new Directed Evolution (DE) techniques it has been possible to improve the performance of existing enzymes by introducing mutations in their primary structures, creating variants that will work much better than the WT.[5]

1.2 Enzyme design

As mentioned before, many transformations carried out in industries do not have a natural enzyme able to perform the reactions in high yields and selectivities, for instance, because the active site can sometimes be too narrow to bind the substrate of interest. These limitations have led to enzyme engineering[5] that consists in three steps: i) selection of mutation points, ii) insertion of the mutations, and iii) evaluation of the activity of the new variants. There are three types of design strategies that can be targeted using computational or experimental approaches: i) non-rational, ii) semi-rational and/or iii) rational strategies.

The first one, Directed Evolution, the non-rational strategy also known as random mutagenesis, is based on generating a huge number of variants evolved in the experimental laboratory inserting random mutations in the enzyme structure. So, the result will be new variants that are active but without knowing the reasons for the enhanced activity. The second one, the semi-rational approach,[6] is a combination of the generation of a series of variants predicted by means of calculations, where later Directed Evolution (DE) will be applied in the laboratory to improve the activity. And finally, the rational strategy consists to limit the number of inserted mutations to a small set of rationally selected ones. This can be achieved, for instance, by using different types of computational programs to understand the effect of the new introduced mutations at the atomistic level. This would provide a more accurate view and understanding of the basis of the activity improvements. And so, by means of computational methods one could answer questions like: How mutations modify enzyme chemo-, regio-, or stereo-selectivities?[5] or how the binding of the substrate is affecting or inducing a change in the conformation of the enzyme?

1.3 Epoxide Hydrolase

From the whole classification of enzymes, we are interested in hydrolases, specifically in epoxide hydrolases (EH). This kind of enzymes can perform asymmetric epoxide hydrations in an enantioselective[7] way to form the corresponding diol products.
It has been long debated the EHs mechanism, but nowadays it is known that it takes place *via* a two-step mechanism,[8] where Tyr144 and 203 residues are responsible for substrate recognition and Asp97, from the catalytic triad (Asp97-His267-Asp239), is the responsible to perform the nucleophilic attack (see in Figure 1 *Bm*EH active site residues). EH from *Bacillus megaterium* (*Bm*EH) catalyses the hydrolysis of terminal epoxides like phenyl or naphthyl glycidyl ether (PGE and NGE). *Bm*EH preferentially opens the epoxide group of the *(R)*

enantiomer of PGE substrate, but with low activities towards the bulky NGE substrate. However, in a work reported by Kong *et al.*[4] the efficiency of *Bm*EH was improved by mutating residues of the substrate-binding site. In addition, Kong *et al.* on the basis of an X-ray structure (PDB: 4O08) identified an active site tunnel for *Bm*EH enzyme and defined one region as the substrate entrance (as zone 1), active site (as zone 2), and product release (as zone 3) (see Figure 1).

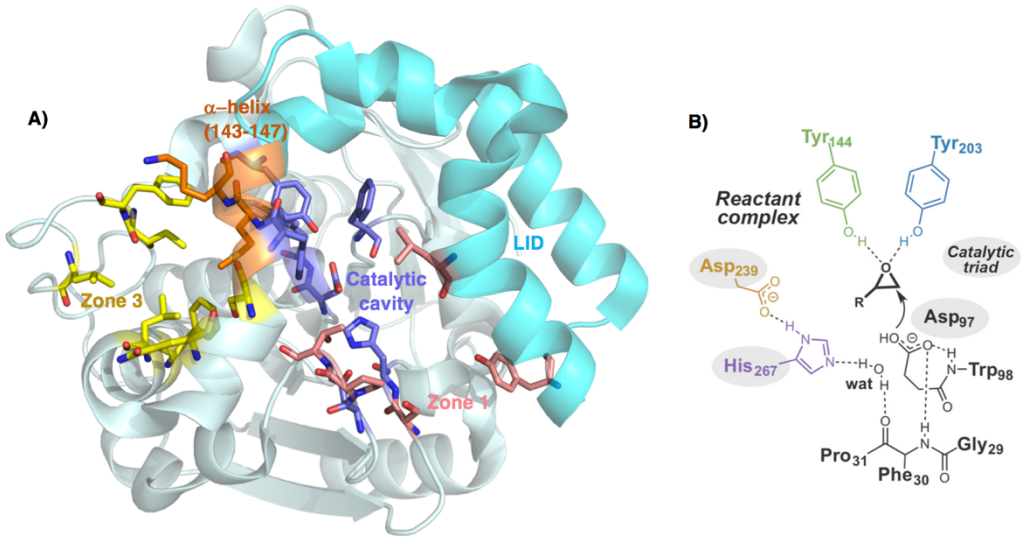


Figure 1. A) Representation of the active site machinery of *Bm*EH enzyme. Residues from the catalytic cavity represented in purple sticks. The *Bm*EH zone 1 (substrate entrance) and zone 3 (product release) sticks are coloured in pink and yellow, respectively, also lid domain is highlighted in cyan and α–helix that contains Tyr144 binding residue in orange. B) Schematic representation of the first step of the reaction.[9]

In a recent computational work[9] from our group the dynamic behaviour of *Bm*EH enzyme has been studied, in which four different conformations for the *Bm*EH WT structure in the *apo* state (without the presence of any ligand) were identified. These four conformational states exist in dynamic equilibrium. In the first conformation, although the catalytic triad and tyrosine residues are well positioned to carry out the catalysis,[9] the volume of the active site is too small, and thus the space for binding the substrate is not enough. Between the first and second conformational states, there is very little difference at a structural level, since only the side chain of some active site residues are slightly different. More interesting, the third and fourth conformational states are clearly different compared to the X-ray like first conformational state. In the third one, the lid that covers the active site is deviated from the first conformation. Apart from that, there are two other structural changes: i) the loop that contains His267 is displaced but remains at a short distance from Asp97, and ii) the loop containing Asp239 adopts a partially open conformation, causing an increase of the volume of the active site.[9] Finally, the fourth conformation, with a completely different structure as

compared to the X-ray like conformation, presents the loop that contains Asp239 completely open. This generates an active site cavity with much wider volume for substrate binding but with the active site residues in a catalytically incompetent orientation for the reaction to take place. For instance, Tyr144 and 203 are also bad positioned to activate the substrate as they point outside the active site. All these conformational changes observed in the *apo* state of the WT enzyme suggested that Asp239 and Tyr144 conformational changes might play an important role for binding the PGE epoxide substrate. Therefore, in this project, Molecular Dynamics (MD) simulations have been performed to study the substrate binding affinity of each conformation with both PGE enantiomers.

2. OBJECTIVES

Enzymes are dynamic biomolecules that are able to explore many different conformations. The huge number of different structures that enzymes can adopt, as well as their probabilities to be explored, can somehow be tuned by the presence of molecules (i.e. population shift), like a substrate or a product, or indeed by the insertion of mutations in the enzyme sequence.[10] All conformations that enzymes can visit in solution might be relevant for their function, but also for processes such as substrate binding or product release. Therefore, Molecular Dynamic (MD) simulations have recently become a very important computational tool in order to accurately understand the dynamic behaviour of the enzymes along different steps in the catalytic cycle.

In a previous computational work in our group,[9] four main conformational states were explored in the absence of any ligand for the *Bm*EH wild-type (WT) enzyme. Two of these states were speculated to be relevant for substrate binding. Therefore, the main goal of this project arises from this hypothesis: perform MD simulations to understand the dynamic behaviour of those four main conformational states for the *Bm*EH WT enzyme, but in the presence of both enantiomers of the natural substrate PGE.

In order to study these enzyme-substrate (E-S) systems and accomplish the main goal, we propose to answer three different questions:

1- How PGE substrate (*R* and *S* enantiomer) is bound into the active site of the EH in the four conformational states?

2- Perform MD for all systems (see Figure 2 in Methodology) and identify if in some of the conformations the substrate is kept in the active site. If this is so, is the substrate PGE found in a catalytically competent binding pose?

3- As mentioned previously, conformational changes were observed to be relevant for substrate binding. Can we observe the closure of the conformation of the loop Asp239 to achieve a well-preorganized X-ray like active site?

The answer to these questions will elucidate which is the best conformation for substrate binding and recognition, as well as for catalysis.

3. METHODOLOGY

In order to perform this project, the following computational tools have been used:

3.1 Molecular docking using Autodock. Before performing Molecular Dynamic (MD) simulations, since we are interested in studying the enzyme-substrate (E-S) complexes (see Figure 2) (in this work PGE natural substrate inside the active site of *Bm*EH WT structure), the substrate has to be docked in the four different conformations of the enzyme under study.

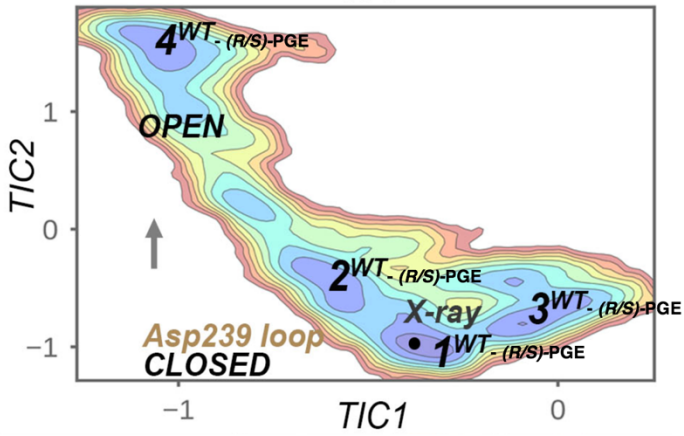


Figure 2. Representation of the free energy landscape where the four most stable conformational states of *Bm*EH enzyme with substrate *(R/S)*-PGE (E-S systems) are represented..

For this, AutoDock,[11] a suite of automated docking tools is used. This software is designed to predict how molecules, such as substrates, bind to a receptor of known 3D structure. In general terms, AutoDock consists of two main programs: i) AutoDock performs the docking of the ligand to a set of grids describing the target protein, and ii) AutoGrid pre-calculates these grids. In more detail, to perform these molecular dockings, a box of 15 Å is defined, whose centre is determined using the coordinates of three atoms of the catalytic residues (OD1 Asp97, OH Tyr144 and 203) (see Figure 3). Along the prediction of the best preferred substrate orientations in the specified region, the natural substrate PGE is allowed to rotate freely (complete flexibility), while the enzyme is modelled as a rigid entity. Binding poses were selected according to the best binding energy scores and by visual inspection with PyMOL.[12]

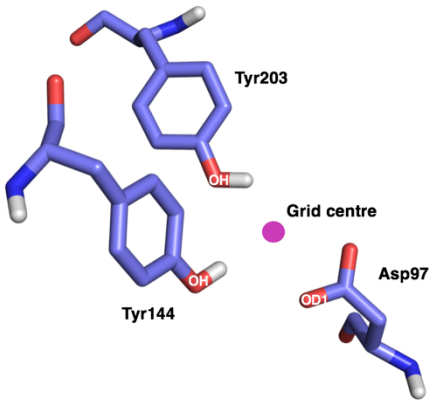


Figure 3. Catalytic residues used for defining the grid centre coordinates of the active-site molecular docking.

3.2 MD simulations of *Bm*EH-PGE systems. Each E-S system (see Figure 2) was immersed in a truncated cubic box with a 10 Å buffer of water molecules but keeping also crystallographic waters. Systems were neutralised by adding explicit counter ions (Na^+ and Cl^-). All subsequent simulations were performed using Amber 99 force field (ff99SB)[13] and the general Amber force field (GAFF)[14] to parametrise the substrate. A two-stage geometry optimisation approach was performed. The first stage minimises only the positions of solvent molecules and ions, and the second stage is an unrestrained minimisation of all the atoms in the simulation cell. The systems were then heated by increasing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions. At the equilibration process, the substrate in most of the cases is displaced from the active site. To solve this problem, a new equilibration strategy has been designed to ensure that, once this step is finished, the substrate remains in the active site at the start of the run production trajectories.

This new strategy is based on stablishing two artificial bonds: i) between the oxygen atom of the epoxide and the hydroxyl group of Tyr203, and ii) between the terminal carbon atom of the epoxide and the OD1 atom of Asp97 (see Figure 3), during the equilibration process to stabilise the E-S structure. The equilibration with these harmonic restraints is performed in a series of 6 steps in which the strength of these artificial bonds is progressively reduced (from a value of a force constant of 80 to 5 $kcal \cdot mol^{-1} \cdot Å^2$). Afterwards, a final equilibration step is done with no restraints, and we are able to start the production run from a good E-S point to explore the behaviour of the E-S system without imposing any artificial situation.

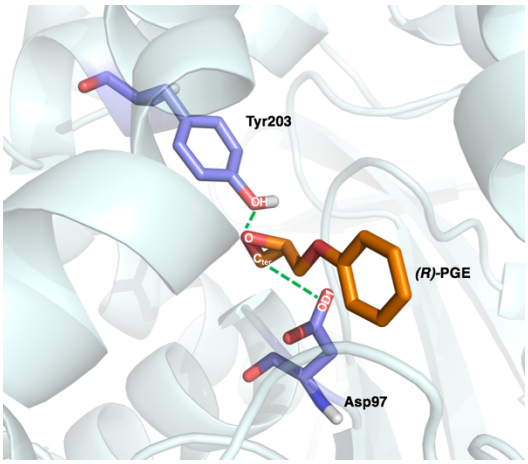


Figure 4. Representation of the restrained bonds in a discontinuous line coloured in green.

3.3 MD data analysis. In order to proceed with the analysis of the MD simulations, the module CPPTRAJ,[15] a software for the analysis and processing of the MD trajectory data is used.

First, it was used to visualise the MD trajectories. The software generates pdb files from the trajectory of the MD simulation file and can be displayed in PyMOL as a movie, so that the conformational dynamics of the enzyme can be visualised during the simulation.

Afterwards, the root mean square deviation (RMSd) (eqn (1)) of the substrate was also computed to be able to study the deviation of the substrate from the starting position along the MD run. This calculation measures the deviation of a set of coordinates object of study with respect to a set of reference coordinates, with RMSd=0, taken from the parameter/topology file of each E-S system. Therefore, a value equal to 0 will indicate that there is no deviation from the set of coordinates studied with respect to the one of reference. If the value of the RMSd is large, it will indicate that there is a substantial deviation from the initial conformation.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2} \qquad (1)$$

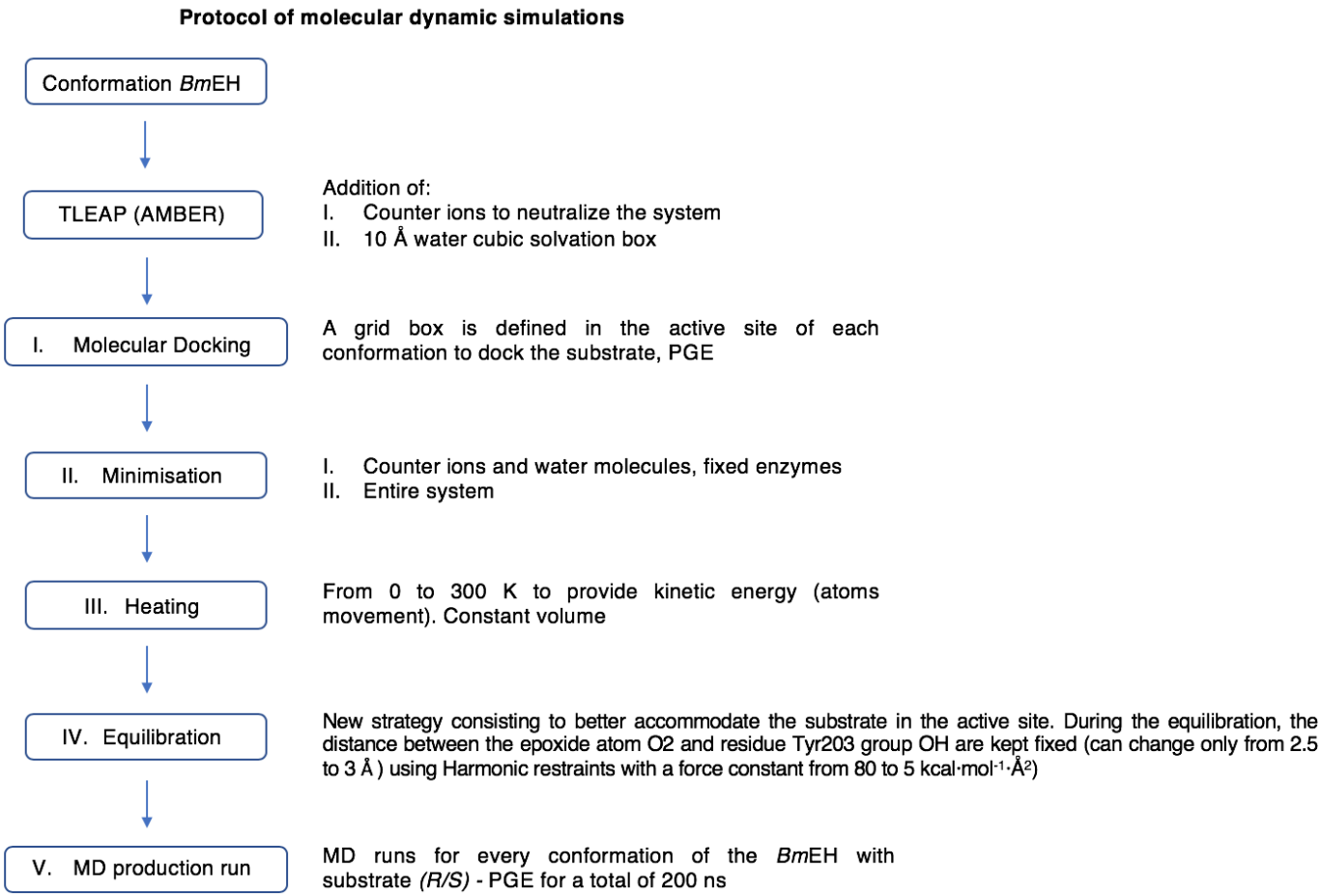
where δ_i is the distance between atom *i* and either a reference structure or the mean position of the *N* equivalent atoms.

Also, the root mean square fluctuation (RMSf) (eqn (2)) of the enzyme was computed in order to identify which were the most flexible regions of the enzyme.

$$RMSF = \sqrt{\frac{1}{T}\sum_{t_j=1}^{T}(x_i(t_j) - x_i)^2} \quad (2)$$

where $x_i(t_j)$ is the coordinates of atom x_i at time t_j, and x_i the reference coordinates.

3.4 Ethical and sustainability. In this work, only computational tools have been used, so no waste is produced. But a lot of electricity is used, since we have used computers to run the calculations and air conditioning to keep computers at a good temperature so that they can work properly. By working with enzymes through computational tools, we reduce a lot of steps for the pharmaceutical and fine industry, as like the synthetic and purification ones, also reducing a large amount of energy and expensive equipment needed and thereby the processes can be carried out at a normal temperatures and pressure.
In addition, data submitted in this work has not been copied or changed in order to achieve better results also, plagiarism has been avoided. All the bibliography used has been listed and detailed in the corresponding section.



Scheme 1. Computational protocol used to explore the binding affinity of the natural phenyl glycidyl ether (PGE) of the epoxide hydrolase (EH) conformational states.

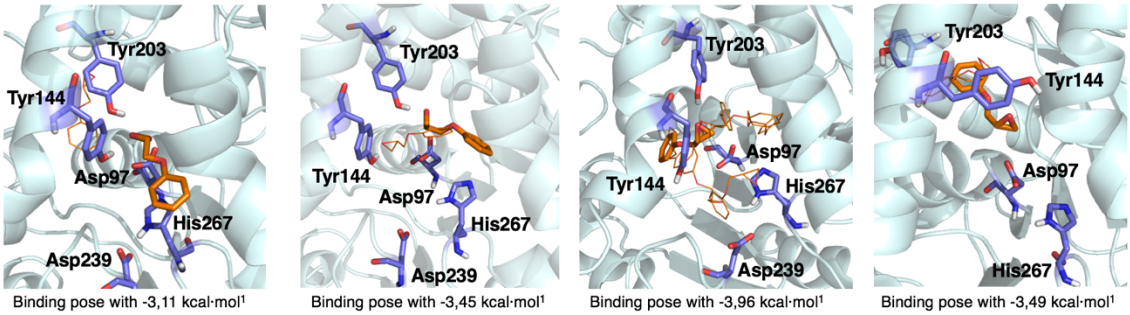
4. RESULTS

As we know, our *Bm*EH enzyme explore four different conformations in the absence of substrate (see Figure 2), where two of these conformations show significant conformational changes for substrate binding. In order to study the E-S systems, that is, the enzyme with the substrate in the active site, the models for each system must be constructed. Therefore, molecular docking is used to explore different positions of the substrate for each conformation, limiting the region of the active site.

4.1 Building the enzyme substrate (E-S) structure

Molecular docking was used to explore the substrate binding pose in the active site of each conformation of the *Bm*EH WT. And so, we defined a box of 15 Å (more details described in Methodology) that includes the catalytic and binding residues in order to explore the preferred orientations of each phenyl glycidyl ether (PGE) enantiomer in the different conformations of the enzyme (1^{st} – 4^{th}). It should be mentioned, that the region that we defined to perform the molecular docking also includes the substrate entrance (zone 1) and the product release (zone 3). Once the molecular docking is finished (see Figure 5) for each *Bm*EH WT conformation with the *(R/S)*-PGE substrate, the different binding poses generated were analysed by visual inspection using PyMOL. In Figure 5 the active site of each enzyme conformation together with all the predicted orientations (in lines) and the selected poses for the *(R)* and *(S)*-PGE substrate (in sticks) according to the best binding energy of the molecular docking and the best substrate orientation for catalysis is shown. As it can be seen in the figure below, in the first and second conformation the docking predicted that the best orientation of both enantiomers was when the epoxide interacts with Tyr144 and 203, most likely due to the low volume of the catalytic cavity and the proper organisation of the catalytic residues. However, in the third and fourth conformation it is observed that the substrate *(R/S)*-PGE is in the active site with the epoxide oriented towards the catalytic residues. Since these are the two conformations that have more volume[9], the substrates have been able to better accommodate inside the active site.

A) *Bm*EH conformations (1st – 4th) with *(R)*-PGE



B) *Bm*EH conformations (1st – 4th) with *(S)*-PGE

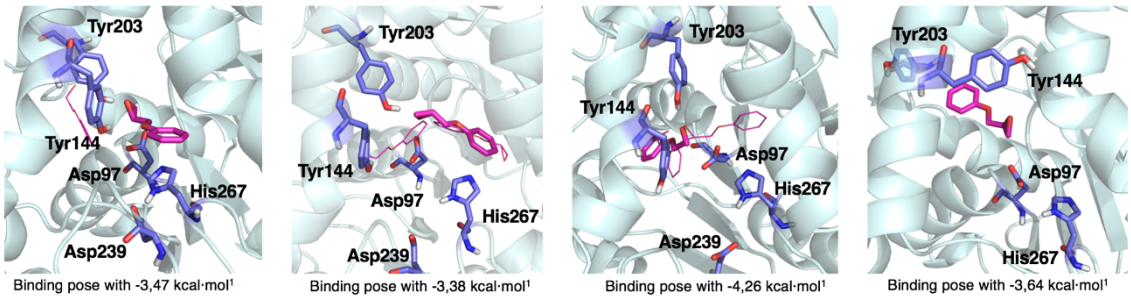


Figure 5. A) *Bm*EH conformations for substrate *(R)*-PGE coloured in orange, and B) *Bm*EH conformations for substrate *(S)*-PGE coloured in magenta. From left to right: 1st, 2nd, 3rd, 4th conformations of the enzyme. Represented in orange sticks best binding pose predicted of the substrate for Autodock and in lines the other (less favourable) binding poses.

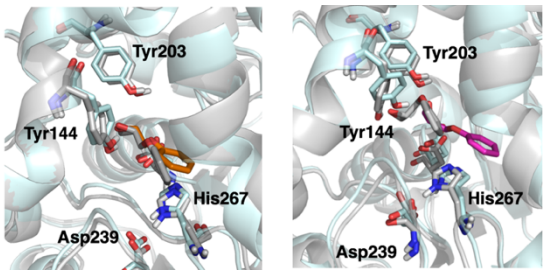
4.2 Molecular dynamic (MD) simulations. Equilibration.

As discussed above (see section 3.2 of Methodology), constructing the E-S model for each system is complicated. The fact that the interaction between Tyr144 (responsible for substrate recognition) and the epoxide is not obtained with docking calculations causes that once the equilibration step has finished the substrate might have left the active site. For this reason, a new equilibration protocol has been designed which allows us to accommodate the E-S systems for the MD simulations.

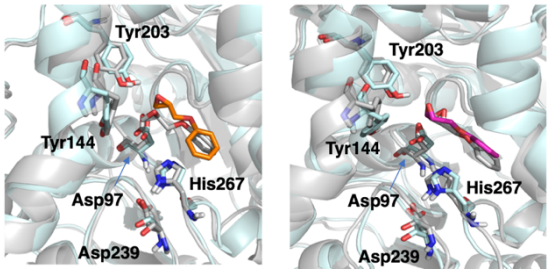
After selecting the best binding pose and building the initial structure for the E-S systems, the equilibration step of the MD simulations was run, which is before production run (see Scheme 1 from Methodology). At that point it was analysed if during the new equilibration protocol, the substrate had been able to better accommodate and remain inside the active site in the last steps in which the restraints are not applied. For this, it has been superimposed the first frame and the last one (see Figure 6) from the whole equilibration step for each E-S system.

As it can be seen, the new equilibration protocol has fulfilled our purpose in most of the cases, because despite not having a good orientation for all cases at least the substrate has been kept well accommodated in the active site of the enzyme. And so, at this point of the MD simulation the E-S system for each case has been properly described better to study the interaction of *(R/S)*-PGE with the catalytic residues.

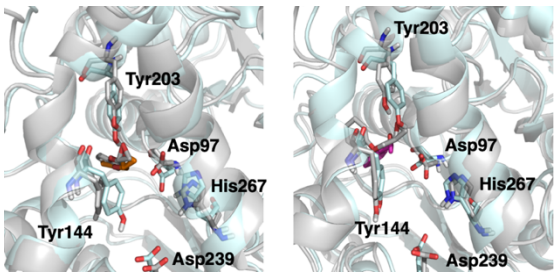
A) *Bm*EH conf. 1 – *(R)*-PGE B) *Bm*EH conf. 1 – *(S)*-PGE



A) *Bm*EH conf. 2 – *(R)*-PGE B) *Bm*EH conf. 2 – *(S)*-PGE



A) *Bm*EH conf. 3 – *(R)*-PGE B) *Bm*EH conf. 3 – *(S)*-PGE



A) *Bm*EH conf. 4 – *(R)*-PGE B) *Bm*EH conf. 4 – *(S)*-PGE

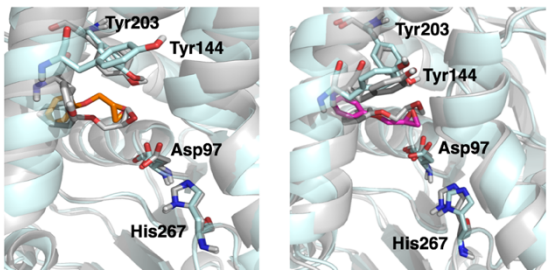


Figure 6. Superimposed structure, of the A) *Bm*EH conf. (1-4) – *(R)*-PGE system and B) *Bm*EH conf. (1-4) – *(S)*-PGE system, of the first and last frame of the new equilibration protocol. Structure coloured in palecyan

corresponds to the start of the equilibration with substrate coloured in orange *(R)* enantiomer and in pink *(S)* enantiomer, and structure coloured in grey corresponds to the end of this step protocol.

4.3 Molecular dynamic simulations. Production run.

After performing the MD simulations for each E-S system (see Figure 2 from Methodology), different analyses were performed to initially identify if the substrate was interacting with the active site along the 200 ns MD simulation (graphs of catalytic distances) and to study the most relevant conformational changes of the enzyme (Root Mean Square deviation (RMSd) and Root Mean Square fluctuation (RMSf)). Additionally, the MD trajectories were also visualised with PyMOL extracting some representative snapshots to deeper analyse the conformational changes. On one hand, this will help to understand the behaviour of each system. So, with this, we would be able to study which is the best conformation to keep the substrate in a binding pose inside the active, but moreover, in which of those good poses the substrate is oriented for a good catalytically situation.

In order to analyse the results of the MD simulations, a comparison between both substrates will be made for each conformation. In this way, we will be able to know which similarities and differences exist between *(R/S)*-PGE enantiomers.

4.3.1 First conformational state of *Bm*EH – *(R/S)*-PGE

This conformation corresponds to the X-ray like conformation[9] and has a well-positioned catalytic triad (Asp97-His267-Asp239) and binding (Tyr144 and 203) residues (see Figure 7) to perform the catalysis.

A) *Bm*EH conf. 1 – *(R)*-PGE B) *Bm*EH conf. 1 – *(S)*-PGE

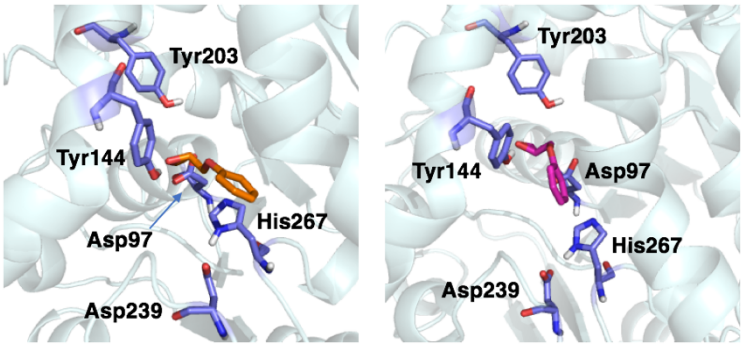


Figure 7. Representation of the substrate binding pose at the start of the MD production run. Left: *(R)*-PGE enantiomer and right *(S)*-PGE enantiomer.

First, it was plotted the measure of the RMSd (see Figure 8) considering only the substrate *(R/S)*-PGE along the 200 ns of the MD simulation to study its deviation from the initial position where it is bound as it was a good orientation for catalysis. This analysis allows us to identify important changes in the mobility in relation to the initial substrate binding pose of the production run simulation. As it can be seen in Figure 8, these values differ a lot from the initial structure, where *(R)* and *(S)* enantiomers have a good starting point, which means that the substrate explores a large number of binding poses different from the starting one.

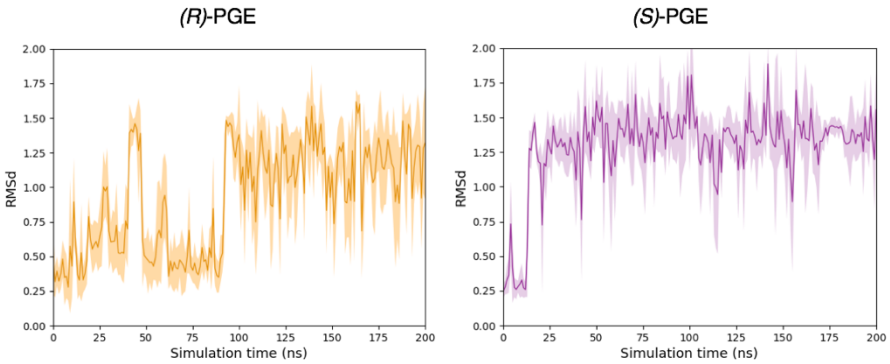


Figure 8. Representation of the RMSd along the MD simulations. Left: *(R)*-PGE enantiomer and right *(S)*- PGE enantiomer.

For both enantiomers, the RMSd values differs a lot from the starting point, this is somehow telling us that the substrate gets out from the active site. We expected that the substrate would be kept in the active site and the catalytic residues well positioned to carry out the catalysis and so, the value of the RMSd would have been lower.

In order to better understand the plot of RMSd, superimposed frames of the MD simulation are shown in Figure 9.

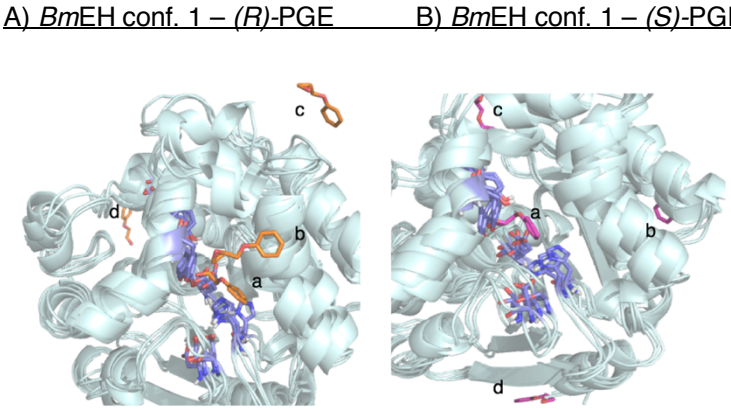


Figure 9. Superimposed structures of 4 frames (a, b, c and d, being 'a' the starting point of the MD simulation and 'd' the last frame). Left: *(R)*-PGE enantiomer and right *(S)*-PGE enantiomer.

This analysis allows us to elucidate how favourable is the interaction between the substrate and the enzyme along the MD simulations and the conformational changes of the substrate and the enzyme. As it can be seen in Figure 9, both *(R/S)*-PGE substrates get out of the active site during the 200 ns. Broadly speaking, there is not a specific conformational change, as for example the opening or closure of the lid, which causes the substrate to emerge from the enzyme. But, if we observe in detail, there are small changes in the position of the catalytic residues. As this conformation has a well-positioned catalytic triad, there must be some interaction between them and the epoxide that will later be lost and hence the substrate leaves the active site of the enzyme. This interaction occurs between the oxygen atom of the epoxide ring and the oxygen atom of Tyr144, therefore it is important for substrate binding although it is weak.

Below are the RMSf graphs (see Figure 10) that allows us to identify, in a local level, which regions of the enzyme are more flexible during the MD simulations. If we focus on the *(R)*-PGE (see Figure 9A), it can be seen that the residues that are from the lid domain (residues from 145 to 202) are those with a greater mobility on the enzyme, because it is the area from which the substrate is bound at the start of the simulation. Therefore, can explain why the substrate leaves the active site, considering that Tyr144 is responsible of substrate recognition. On the other hand, for the *(S)* enantiomer, there is also quite a lot of mobility in the same region, but a large fluctuation from the catalytic residue Asp239 is clearly seen. The substrate starts the MD simulation from the active site (see Figure 9B) and goes through zone 1 (substrate entrance), where Asp239 is found. This explains why this residue has a more marked fluctuation, i.e. to allow the substrate *(S)*-PGE to leave the active site. Thus, the substrate leaves the catalytic cavity just at the start of the MD simulation.

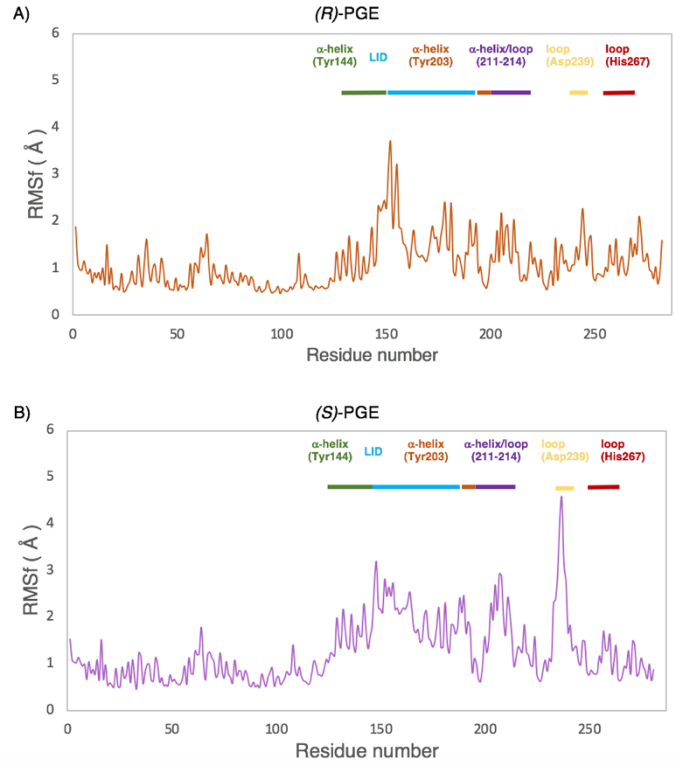


Figure 10. Representation of the RMSf along the MD simulations. A) *(R)*-PGE enantiomer and B) *(S)*-PGE enantiomer.

Altogether, these simulations indicate that conf. state 1 has a weak affinity towards both *(R)* and *(S)*-PGE substrates although the catalytic residues are well positioned.

4.3.2 Second conformational state of *Bm*EH – *(R/S)*-PGE

This conformational state has only some side chain residues slightly different in respect from the first conformation. In the figure below (see Figure 11), it is observed that *(R)* enantiomer starts from a favourable position so that the catalysis can take place, since the epoxide ring is close to Asp97. On the other hand, the starting point of *(S)* enantiomer is in zone 1 (substrate entrance).

A) *Bm*EH conf. 2 – *(R)*-PGE B) *Bm*EH conf. 2 – *(S)*-PGE

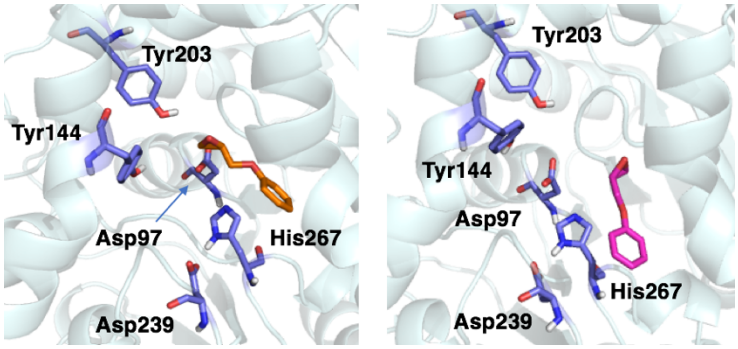


Figure 11. Representation of the substrate binding pose at the start of the MD production run. Left: *(R)*-PGE enantiomer and right *(S)*-PGE enantiomer.

For this conformation, it is observed that the value of the RMSd (see Figure 12), takes a rather high value for both enantiomers. A detailed visualisation of the MD simulations indicates that there is a strong interaction between the oxygen of *(R)*-PGE substrate with Tyr144, which does not happen in the first conformational state. But due to a slight change in the arrangement of the residues that make up the alpha helix, the substrate stops interacting with this Tyr144 and leaves the active site through zone 1 (see Figure 13). Interestingly the case of *(R)*-PGE enantiomer remains close to region 1 (substrate entrance, detailed in Figure 1). However, the *(S)* enantiomer can be found first ahead of zone 1 interacting with one residue of this area, and after this interaction stops it can be found around the surface of the enzyme in the solvent region.

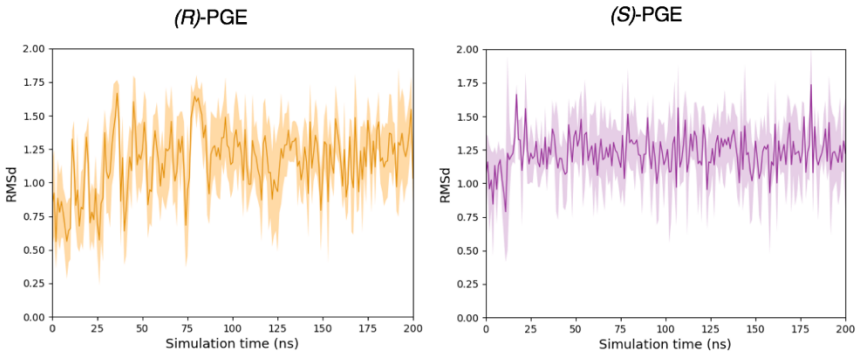


Figure 12. Representation of the RMSd along the MD simulations. Left: *(R)*-PGE enantiomer and right *(S)*-PGE enantiomer.

A) *Bm*EH conf. 2 – *(R)*-PGE B) *Bm*EH conf. 2 – *(S)*-PGE

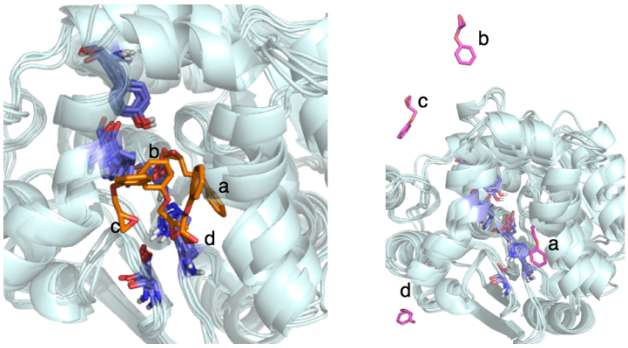


Figure 13. Superimposed structures of 4 frames (a, b, c and d, being 'a' the starting point of the MD simulation and 'd' the last frame). Left: *(R)*-PGE enantiomer and right *(S)*-PGE enantiomer.

If we compare how this conformation of the *Bm*EH enzyme (see Figure 14) fluctuates with respect to the first conformation (see Figure 10), it can be observed that there is a big difference. For this conformation, although the residues that make up the lid domain have some mobility along the 200 ns, there is no residue or region of the enzyme with a high elevated fluctuation value.

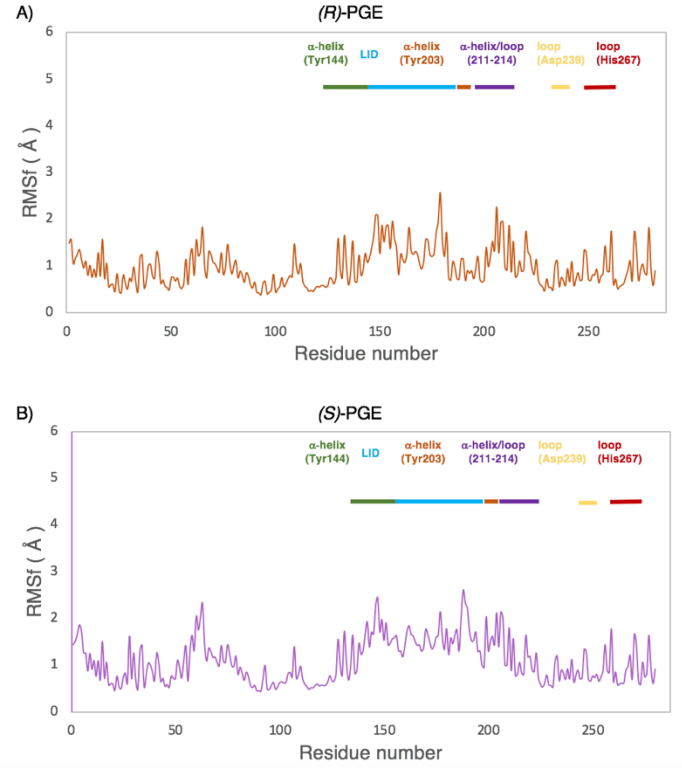


Figure 14. Representation of the RMSf along the MD simulations. A) *(R)*-PGE enantiomer and B) *(S)*-PGE enantiomer.

Our analysis has shown that conf. 2 state has a much more favourable binding affinity towards *(R)*-PGE than *(S)*-PGE. This is in line with the experimentally reported selectivity of *Bm*EH since this enzyme catalyses the opening of the epoxide ring of the *(R)* enantiomer.

4.3.3 Third conformational state of *Bm*EH – (*R*/S) PGE

In this state, conf. 3, there is a great conformational change as the loop that contains His267 (catalytic residue) is a little bit displaced (see Figure 15) but still close to the catalytic Asp97. This causes an increase of the active site volume and, therefore, the substrate can be well accommodated in it. So, it is expected that both enantiomers of PGE substrate will remain inside the active site along the MD simulation but exploring different binding pose orientations with the catalytic residues. Although in this conformation the natural substrate is inside the active site, the epoxide ring is not well oriented towards Tyr144 (substrate recognition) and Asp97 (nucleophilic attack).

A) *Bm*EH conf. 3 – *(R)*-PGE B) *Bm*EH conf. 3 – *(S)*-PGE

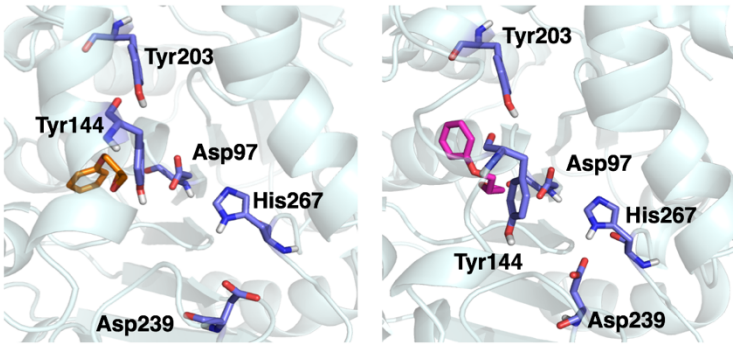


Figure 15. Representation of the substrate binding pose at the start of the MD production run. Left: *(R)*-PGE enantiomer and right *(S)*-PGE enantiomer.

Following the same analysis performed for the other two conformations, the values of the RMSd of the substrate along the 200 ns are represented in the Figure 16.

As this conformation has more volume in its active site (227 $Å^3$)[9], it is expected that the substrate will remain inside the active site exploring a large number of different binding poses and thus the value of the RMSd would be higher.

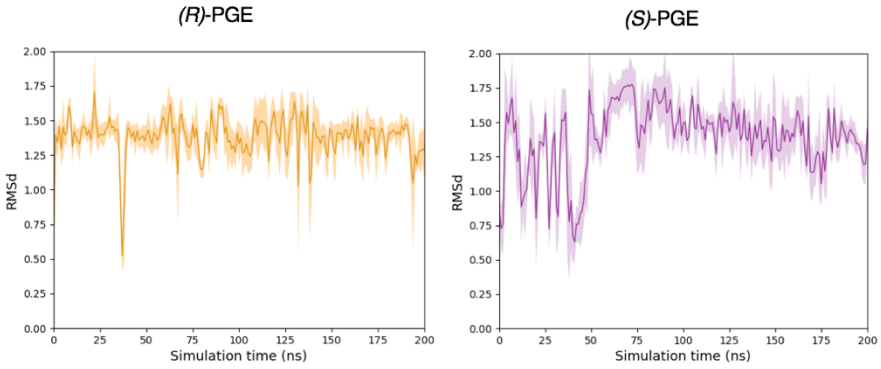


Figure 16. Representation of the RMSd along the MD simulations. Left: *(R)*-PGE enantiomer and right *(S)*- PGE enantiomer.

As it can be seen in the graphs, the *(R)* enantiomer has explored a few binding poses where its deviation value is very lower than 1 (approximately, 0.5). This indicates that the substrate has a fairly good binding pose to carry out the catalysis, since this orientation is close to the initial position of the MD simulation, which is a good competent position for us. In the case of the *(S)*-enantiomer, this explores several competent binding pose orientations since it can be observed that during the first 50 ns of the simulation, there are several peaks with values of deviation of 0.5.

As shown in Figure 17 the substrate remains inside the enzyme specifically in the catalytic cavity. *(R)*-PGE substrate (see Figure 17A) explores an orientation towards the product release area (region 3), while the *(S)*-PGE remain very close to Tyr 144 and 203 (in this system oriented towards region 1), and from the nucleophilic residue Asp97 along the MD simulation. Despite in a work by Kong *et al.*[4] it was reported that the product release area is blocked by Met145 residue, the fact that the substrate *(R)*-PGE is able to orientate towards this area, is due to the slightly open conformation of the active site residues that make more space in this cavity and, for substrate *(S)*-PGE to better accommodated inside it.

A) *Bm*EH conf. 3 – *(R)*-PGE B) *Bm*EH conf. 3 – *(S)*-PGE

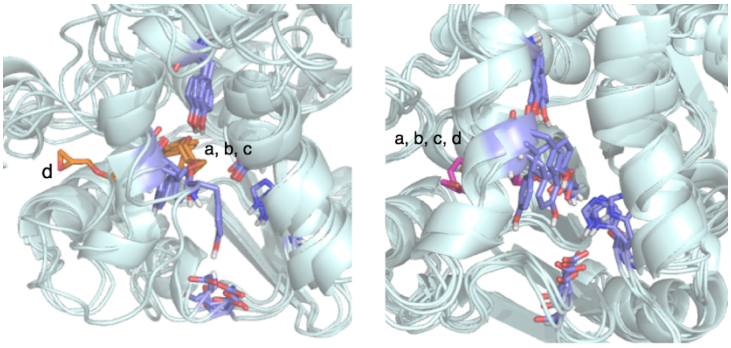


Figure 17. Superimposed structures of 4 frames (a, b, c and d, being 'a' the starting point of the MD simulation and 'd' the last frame). Left: *(R)*-PGE enantiomer and right *(S)*-PGE enantiomer.

In order to better understand the conformational dynamics of this conformation of the *Bm*EH enzyme, we can see that in the graphs of the RMSf (see Figure 18) Tyr144 (active site residues), residues that make up the alpha helix (211-214) and the loop containing residue Asp239 for substrate *(R)*-PGE fluctuates a lot with respect to the rest of the enzyme. As in this conformation the enzyme has a large catalytic cavity, the residues that make up these regions have much higher mobility and so, the substrate can explore different orientations making the E-S system having a good accommodation.

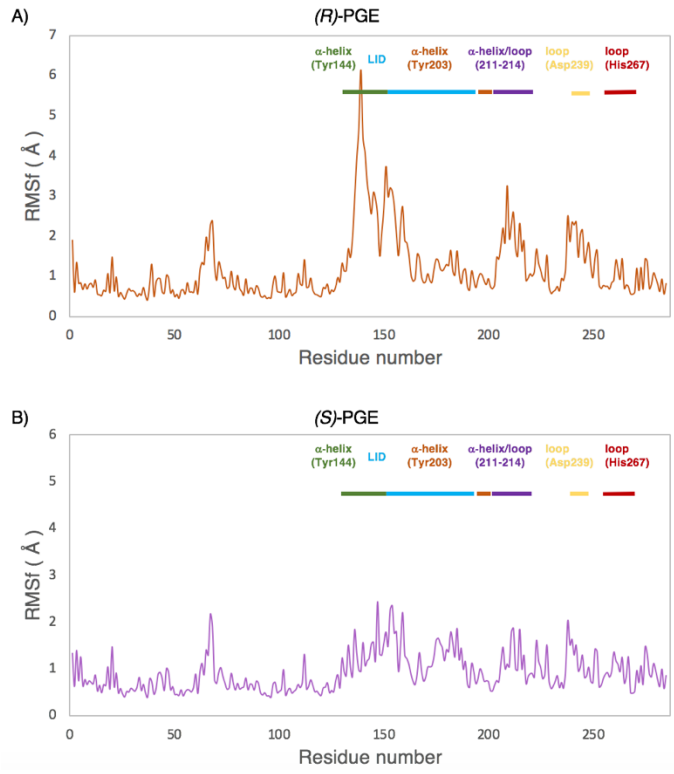


Figure 18. Representation of the RMSf along the MD simulations. A) *(R)*-PGE enantiomer and B) *(S)*-PGE enantiomer.

These analyses provide us that for this conf. 3 state, the *Bm*EH – *(R)*-PGE system, is ideal in order to carry out the catalysis since the catalytic residues and the substrate adopt some competent catalytic conformations along the 200 ns. While the *(S)* enantiomer, despite adopting competent orientations, both Tyr144 and 203 residues are poorly oriented.

4.3.4 Fourth conformational state of *Bm*EH – (*R*/S) PGE

This conformational state 4, presents the Asp239 (see Figure 19) containing loop in a completely open conformation and the lid domain is also displaced.

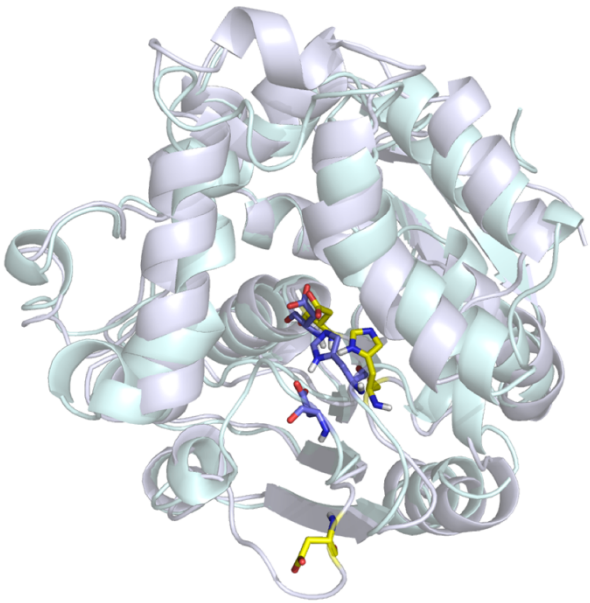


Figure 19. Open conformation of Asp239 containing loop. Cartoon represented in palecyan and blue sticks conformational state 1 and, cartoon coloured in bluewhite and sticks in yellow conformational state 4.

Since this conformation has the biggest active site pocket (617 Å^3)[9], it is also expected that the substrate will remain in its catalytic cavity (as observed in the third conformational E-S system) and that will explore a large number of competent binding poses.

A) *Bm*EH conf. 4 – *(R)*-PGE B) *Bm*EH conf. 4 – *(S)*-PGE

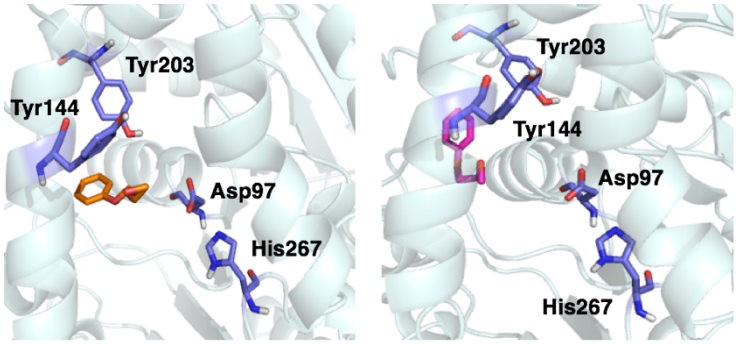


Figure 20. Representation of the substrate binding pose at the start of the MD production run. Left: *(R)*-PGE enantiomer and right *(S)*-PGE enantiomer.

Since this conformation has the largest active site pocket, it is clear that both enantiomers have explored a large number of different binding poses (see Figure 21) despite a bad starting point (see Figure 20). As shown by RMSd many peaks with a deviation value of RMSd close to 0 for *(R)*-enantiomer, whereas for *(S)*-enantiomer less productive poses are explored.

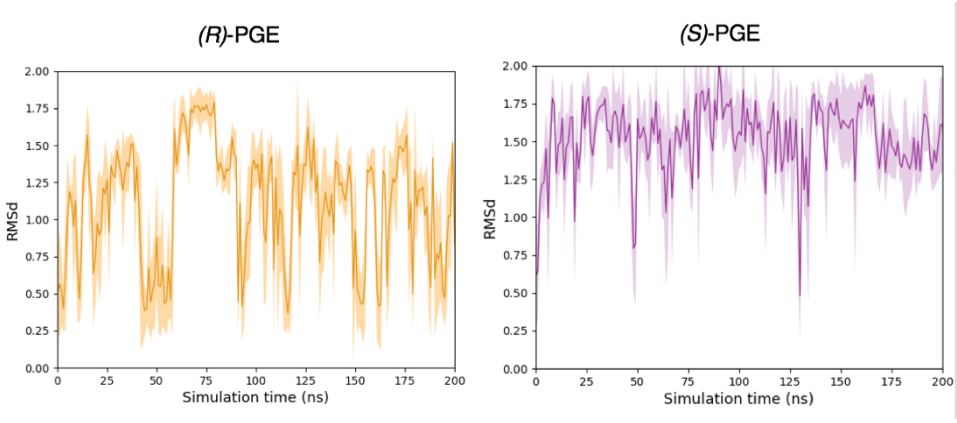


Figure 21. Representation of the RMSd along the MD simulations. Left: *(R)*-PGE enantiomer and right *(S)*- PGE enantiomer.

A careful visualisation of the MD simulations shows a conformational change at the active site (see Figure 22). The superposition of some frames along the 200 ns (see Figure 22), shows that Tyr144 is facing two different orientations (upwards and towards zone 1), both out from the active site, due to the closing of the lid domain. As this residue is responsible of substrate recognition and activation, the epoxide ring opening reaction will be substantially hindered. However, especially for *(R)*-PGE some poses present Tyr144 properly positioned for the recognition and activation. While *(S)*-PGE can be found most of the MD simulation in the upper zone of the lid domain, above the catalytic cavity.

A) *Bm*EH conf. 4 – *(R)*-PGE B) *Bm*EH conf. 4 – *(S)*-PGE

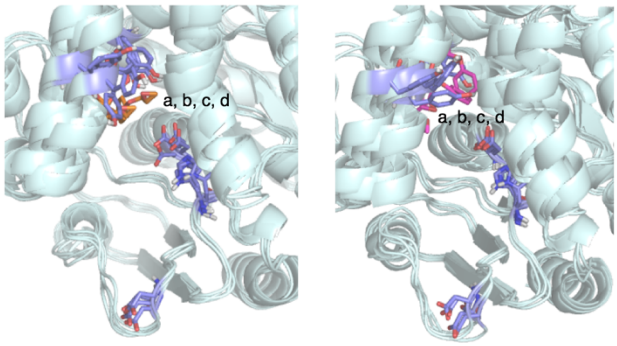


Figure 22. Superimposed structures of 4 frames (a, b, c and d, being 'a' the starting point of the MD simulation and 'd' the last frame). Left: *(R)*-PGE enantiomer and right *(S)*-PGE enantiomer.

As it can be seen in the RMSf graphs (see Figure 23), the regions that have higher mobility of the enzyme is the lid domain and the residues from the loop where is found Asp239. These two local changes in the structure of the enzyme cause the best accommodation of both enantiomers in the active site, thanks to the fact that the cavity has gained volume and, that the catalysis cannot be carried out.

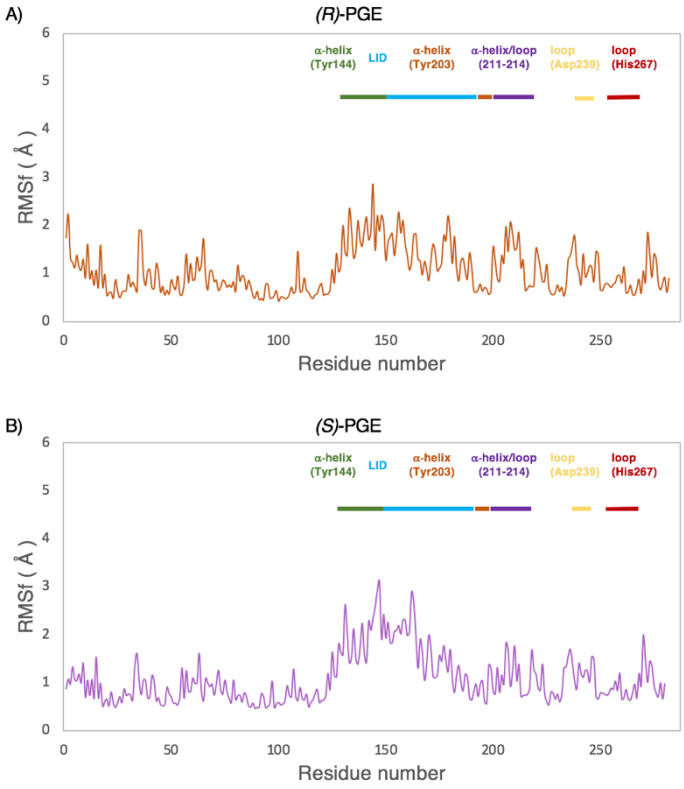


Figure 23. Representation of the RMSf along the MD simulations. A) *(R)*-PGE enantiomer and B) *(S)*-PGE enantiomer.

As expected the substrate *(R/S)*-PGE remains inside the active site but in no case could the catalysis be carried out since the catalytic residues are not well oriented, specially Tyr144 responsible for recognition and activation of the natural substrate.

Broadly speaking it is possible to say that the best conformational state for the reaction to take place is the third one. In the latter, although *(R)*-PGE, which is the enantiomer we are interested in, starts from an unfavourable position, it explores other conformations catalytically competent for catalysis, some of them quite close to the ideal QM geometry. In contrast, *(S)* enantiomer, which also starts from an unfavourable position, is kept in the active site but away from Tyr144 and Asp97. As we have seen, the first conformation has a very small catalytic cavity, and the substrate leaves the active site just at the beginning of the MD simulation. However, both enantiomers present weak interactions with Tyr144 that are lost due to a conformational change of the enzyme. Although in the second conformation a good interaction between *(R/S)*-PGE and the enzyme is well established, the substrate does not adopt any position that approaches the ideal one to carry out the catalysis, and the catalytic cavity is not sufficiently wide to accommodate the substrate inside. On the other hand, it is very interesting that *(R)*-PGE substrate remains ahead of zone 1 interacting weakly with Tyr144. In the fourth conformational state the substrate explores more binding poses that are catalytically competent. However, only in a few of them (specially for *(R)*-PGE) Tyr144 is properly positioned for recognising and activating the epoxide substrate.

5. CONCLUSIONS

In this work, we have studied the *Bm*EH – *(R/S)*-PGE systems in four previously reported conformational states using Molecular Dynamic (MD) simulations. Being the first conformation like the X-ray structure, where the catalytic residues are well positioned to carry out the catalysis and, the fourth conformation, where the structure of the lid domain is completely disordered and the loop containing catalytic residue Asp239 in an open conformation. Therefore, our purpose was to study how the substrate *(R/S)*-PGE is bound and its interaction with the active site residues of these conformations, going from one with a competent catalytic cavity to carry out the reaction, to one that has the catalytic residues completely disordered that makes it difficult for the catalysis to take place but structurally, in terms of volume, interesting because the substrate can bound.

To this aim, the PGE substrate was docked in the four conformations of the *Bm*EH enzyme, the best binding pose prediction for each system was prepared, following Scheme 1 from Methodology, for the MD simulation. Once analysed the results of each system and comparing the similarities and differences between the enantiomers for each conformation, the main conclusions obtained can be summarised as follows:

- Conformation 1, that presents the well positioned catalytic triad, does not have a wide active site for the substrate to remain in the active site.
- It has been shown that for the substrate binding in the catalytic cavity, the loop containing Asp239 has to adopt a slightly open conformation.
- The best system to accommodate the substrate *(R/S)*-PGE into the active site is the third conformational state, as the slightly open conformation of Asp239 loop gives additional volume to the cavity.
- From our MD simulations it has been possible to confirm that *(R)*-PGE substrate is the one that has the best binding affinity for the *Bm*EH enzyme, which is in line with the experimental data.

To further study the relationship between the slightly open conformation of the active site and the catalysis, QM and MD calculations should be done.

6. BIBLIOGRAPHY

(1) Rasor, J. P., Voss, E. (2001) 'Enzyme-catalysed processes in pharmaceutical industry', *Applied Catalysis A: General*, 221, 145-158

(2) Bornscheuer, U.T., Huiman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C., Robins, K. (2012) 'Engineering the third wave of biocatalysis', *Nature*, 485, 185-194

(3) Wandrey, C., Liese, A. and Kihumbu, D. (2000) 'Industrial biocatalysis: Past, Present and Future', *American Chemical Society,* 4(4), 286-290

(4) Kong, X.-D., Yuan, S., Li, L., Chen, S., Xu, J.-H., Zhou, J. (2014) *Proceedings of the National Academy of sciences of the United States of America*, 111, 15717

(5) Romero-Rivera, A., Garcia-Borràs, M., Osuna, S. (2017) 'Computational tools for the evaluation of laboratory-engineered biocatalysis', *Chem. Commun.,* 53, 284

(6) Chica, R. A., Doucet, N. and Pelletier, J. N. (2005) 'Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design', *Current Opinion in biotechnology,* 16(4), 378-384

(7) Kong, X.-D., Yu, H.-L., Yang, S., Zhou, J., Zeng, B.-B., Xu, J.-H. (2015) 'Chemoenzymatic synthesis of (*R*)- and (*S*)-propranolol using an engineered epoxide hydrolase with a high turnover number', *Journal of Molecular Catalysis B: Enzymatic*, 122, 275-281

(8) Serrano-Hervás, E., Garcia-Borràs, M., Osuna, S. (2017) 'Exploring the origins of selectivity in soluble epoxide hydrolase from Bacillus megaterium', *Org. Biomol. Chem. The Royal Society of Chemistry*, 15(41), 8827-8835

(9) Serrano-Hervás, E., Casadevall, G., Garcia-Borràs, M., Feixas, F. and Osuna, S. (2018) 'Epoxide hydrolase conformational heterogeneity for the resolution of bulky pharmacologically-relevant epoxide substrates', *Chem. – Eur. J*

(10) Maria-Solano, M.A., Serrano-Hervás. E., Romero-Rivera, A., Iglesias-Fernández, J., Osuna, S. (2018) 'Role of conformational dynamics in the evolution of novel enzyme function', *Chem. Commun.,* 54 (50), 6622-6634

(11) Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009) 'Autodock 4 and AutoDockTools4: automated docking with selective receptor flexibility', *J. Computational Chemistry*, 16: 2785-91

(12) DeLano, W. L. (2002) 'Pymol: An open-source molecular graphics tool', *CCP4 Newsletter on Protein Crystallography*, 40, 82-92

(13) Wang, J., Cieplak, P., Kollman, P. A. (2000) *Journal of Computational Chemistry*, *21*, 1049

(14) Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P.A., Case, D. A. (2004) 'Development and testing of a general AMBER force field', *Journal of Computational Chemistry,* 25, 1157-1174

(15) Roe, D. R., Cheatham, T. E. (2013) *Journal of Chemical Theory and Computation*, *9*, 3084