Computational study of the hydroxylation and epoxidation by the mutants of the PaDa-I peroxygenase variant

Mireia Martínez Sugrañes

UNIVERSITAT ROVIRA I VIRGILI

FINAL DEGREE PROJECT

Project's tutor: Josep Gómez Alvarez

Professional tutor: Victor Guallar Tasies

Biochemistry and Molecular Biology,

Faculty of Chemistry

University Rovira i Virgili,

Tarragona, September 2021

Project based on the results obtained during my curricular and extracurricular internship in the Electronic and Atomic Protein Modelling (EAPM) group located at the Barcelona Supercomputing Center (BSC). This work has been supervised by of Victor Guallar Tasies.

“Always remember,
your focus determines your reality.”
(Qui-Gon Jinn, Star Wars: Episode I – The Phantom Menace)

Index

1. Abstract

Unspecific peroxygenases (UPOs) (EC 1.11.2.1) represent a relatively new family of fungal heme-thiolate enzymes with high biotechnological interest. Many of these enzymes have been engineered, combining rational and directed evolution with computational design methods, to obtain the required selectivity, efficiency and stability for their applications. In fact, using computational approaches it is now possible to map the protein-ligand energy landscape, being able to reproduce a full binding mechanism or an active site induce fit in a very short time. For this purpose, the Protein Energy Landscape Exploration (PELE) software has been used. Using as departure point the UPO mutant from *Agrocybe aegerita* PaDa-I, we characterized the specificity, regio- and enantioselectivity of thirty mutants coming from PaDa-I with five different substrates. This analysis will enlarge the existing data and it is going to set a starting point for future studies with more substrates to achieve the design of new specific mutants.

All 30 mutants' structures were created from the structure of PaDa-I and were prepared modelling the heme site the two-electron-oxidized compound I. A docking with Schrodinger's Glide was performed placing the ligands at the entrance of the heme channel. After, ligand pose refinement was performed using the enhanced-sampling version of PELE, adaptive-PELE.

Computational results showed the more interesting mutants for every substrate, using the residence time and percentage of accepted steps in different thresholds as a criteria. Substrates *cis*- and *trans*-β-methylstyrene showed similar results because of its almost identical structures as well as styrene and the carbon 1 of ethylbenzene. This later two have the highest values of residence time (~24%) and this could be due to their smaller size. In ethylbenzene's regioselectivity, orto has lower residence times but closer poses and higher energies and the opposite happens with para position. In 1,2,3,4-tetrahydronaphtalene's regioselectivity, the pair C2/C3 have better residence times probably due to the major easiness to be orientated towards the heme. In general, mutants 10, 13 and 23 have obtained good results in all cases. Switching to the enantioselectivity, in *cis*- and *trans*-β-methylstyrene, some mutants inverted the ratio of PaDa-I. In styrene, many mutants were polarized towards the R enantiomer, unlike the reference (50/50). For ethylbenzene, several mutants showed predilection for the R enantiomer as PaDa-I and the contrary happens with 1,2,3,4-tetrahydronaphtalene, where variants show bias for the S enantiomer.

This study explores new insights on mutants derived from PaDa-I. The differences between these and the reference will help to do a retrospective analysis to identify the possibly mutations that module the binding energies, regio- and enantioselectivities in each substrate. It will be possible to perform a prospective rational analysis to combine the best mutations in new variants.

Keywords: *computational chemistry, hydroxylation, epoxidation, regioselectivity, enantioselectivity, Monte Carlo molecular simulations, PELE, adaptive-PELE, fungal unspecific peroxygenases, PaDa-I.*

2. Introduction

2.1. The Unspecific Peroxygenases (UPOs)

Unspecific peroxygenases (UPOs) are heme-thiolate peroxidases with mono(per)-oxygenase activity for the selective oxyfunctionalization of C-H bonds. This type of reaction in organic molecules is one of the most challenging tasks in the field of organic synthesis. Despite of all progress achieved, not all requirements of synthetic applications can be fulfilled yet[1].

To achieve these reactions, typically transition metal catalysts were used to activate the inert C-H bonds. After, due to their catalytic diversity, there were used biological networks as cytochrome P450 monooxygenases. UPOs share most of the catalytic attributes of P450 monooxygenases. Unlike UPOs, P450s possess several drawbacks such as their intracellular situation that entails low stability, moderate turnover numbers and the requirement of expensive electron delivering systems like redox cofactors, and auxiliary flavoproteins[2].

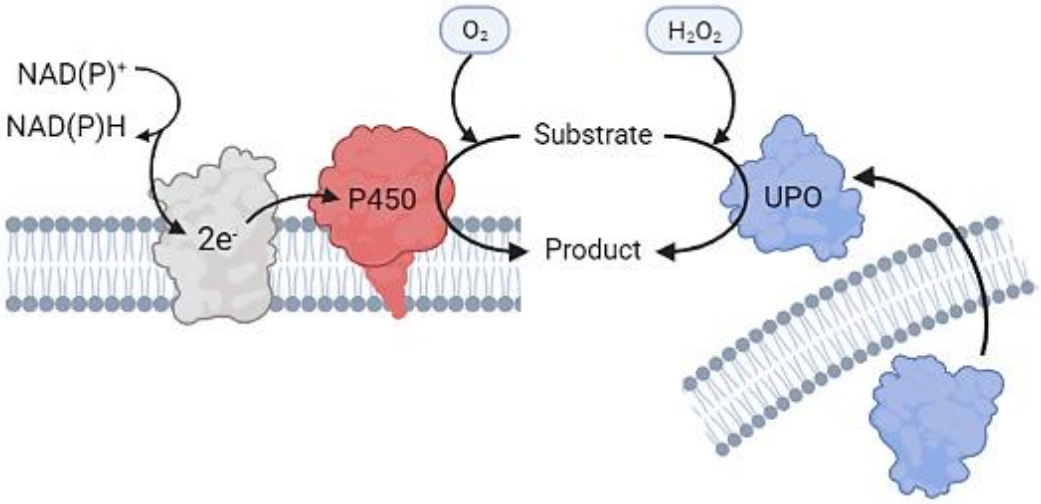


Figure 1. Differences between P450s and UPOs. Intracellular P450s (left) require a source of reducing power (NAD(P)H) and an auxiliary flavin-containing reductase or protein (and often result in futile). Secreted UPO just needs a source of H_2O_2 to be activated (being also more robust due to its extracellular nature).

Meanwhile many steps were taken to overcome the limitations of P450s, the first UPO was identified and isolated from the fungal kingdom[3]. Fueled by catalytic concentrations of H_2O_2, this soluble, extracellular, and very stable enzyme became a potential biocatalyst of great interest in organic synthesis.

2.1.1. Discovery, classification, and distribution

The first unspecific peroxygenase (UPO) was described in 2004[3] as an *Agrocybe aegerita* haloperoxidase for the respective fungus *(syn. Agrocybe cylinracea, Cyclocybe aegerita)*. It belongs to the Basidiomycota phylum (family Strophariaceae) and is commonly known as the Black poplar mushroom. This fungus is found in Europe, North America and Asia and prefers warm and mild climates. It grows on the wood of poplars (*Populus spp.*) and other broad-leaved trees (also known as hardwoods) and origins a white rot. This edible mushroom is popular in some Mediterranean countries, especially in Italy.

The first article didn't use the term peroxygenase because they were focused on the ability of the enzyme to oxidize halides and aryl alcohols[3]. Later, its unique oxygen atom transfer potential was recognized by the hydroxylation of naphthalene[4]. Over the next years, additional aromatic, heterocyclic and aliphatic substrates were found to be subjects of peroxygenation[5–7] and the name of the enzyme changed from haloperoxidase[3] to haloperoxidase-peroxygenase[8]. Later, due to the identification of similar biocatalysts in other fungi, the name was changed to *Agrocybe aegerita* aromatic peroxygenase (APO)[9] following the term "aromatic monooxygenases" used for some P450 enzymes. Finally the name was evolved to unspecific peroxygenase (UPO)[10]. Due to the discovery of many more unspecific peroxygenases, they are systematically abbreviated by the capital letter of the genus and the first and second letter of the species name plus the acronym UPO (in this case *Aae*UPO)[11]. The term "unspecific" comes from their chimeric properties (putting together the catalytic attributes of P450s, classical heme-peroxidases, and chloroperoxidases (CPOs)).

Even more, UPOs are also classified as a member of the heme-thiolate peroxidases (HTP)[12] superfamily due to their characteristic heme ligation by a cysteine and their relation to CPOs.

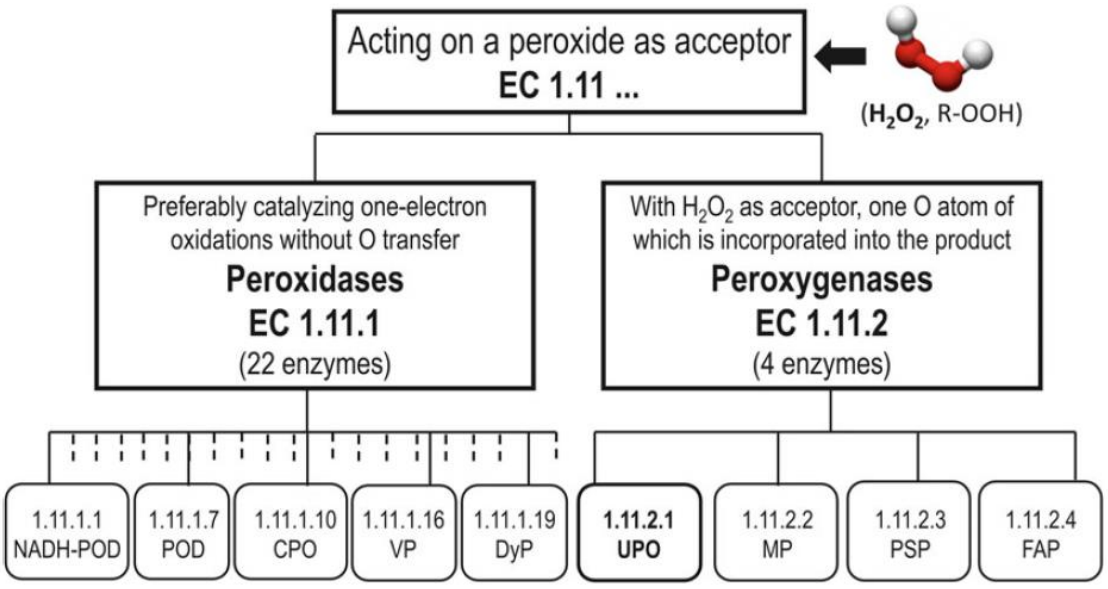


Figure 2. Classification of enzymes using peroxide as the electron acceptor (EC 1.11; peroxidases and peroxygenases) according to the enzyme nomenclature system. ***NADH-POD*** *NADH peroxidase,* ***POD*** *peroxidase (phenol oxidizing),* ***CPO*** *chloroperoxidase,* ***VP*** *versatile peroxidase,* ***DyP*** *dye decolorizing peroxidase,* ***UPO*** *unspecific peroxygenase,* ***MP*** *myeloperoxidase,* ***PSP*** *plant seed peroxygenase,* ***FAP*** *fatty acid peroxygenase. Extracted from Hofrichter, M., Kellner, H., et al. (2015) Fungal Unspecific Peroxygenases: Heme-Thiolate Proteins That Combine Peroxidase and Cytochrome P450 Properties. In Hrycay EG and Bandiera SM, editors, Monooxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome P450 (341-361). Springer International Publishing Switzerland.*

Through a hybrid catalytic mechanism, UPOs show both peroxidative (one-electron oxidation) and, more notably, peroxygenative activity (oxygen-transferring two-electron oxidation with peroxide as the oxygen source)[13]. The latter is the one that is responsible of UPOs' promiscuity in a wide portfolio of transformations. Peroxygenase activities allude to the transfer of a peroxide-borne oxygen atom to substrates. Biocatalysts that catalyze these reactions are classified in a separate sub-class, EC 1.11.2, in the enzyme

nomenclature system [www.chem.qmul.ac.uk/iubmb/enzyme/EC1/11/2/]. The sub-subclass was approved in February 2011 and currently comprises four members (as can be seen in *Figure 2*), among which the unspecific peroxygenase (UPO, EC 1.11.2.1) is the most prominent because of its frequency in fungal organisms and promiscuity for oxygen transfer reactions.

The peroxygenase reaction can be illustrated in a simplified form as shown in the equation below where AH is the substrate, ROOH represents the hydroperoxide, R is an organic substituent or hydrogen atom, AOH designates the hydroxylated product and ROH depicts the reduced hydroperoxide or H_2O.

$$A\text{–}H + ROOH \rightarrow A\text{–}OH + R\text{–}OH$$

The three well-studied UPOs are *Aae*UPO *(Agrocybe aegerita)*, *Cra*UPO *(Coprinellus radians)* and *Mro*UPO *(Marasmius rotula)*[14]*.* All of them have differences with respect to the oxidation of different types of compounds. Besides these three UPO producers, here are many other mushroom species that secret UPOs (for example, *Agaricus bisporus, Mycena galopus, Serpula lacrymans)*[11]*.* This information of the occurrence of UPO-like enzymes can be obtained from genetic databases where approximately 2.000 sequences of presumably UPO enzymes are found. *Figure 3* shows this diversity using a phylogenetic tree of UPOs covering 30 representative and very diverse fungal species. However, as can be seen in the *Figure 3,* there is evidence of the presence of UPO genes outside the true fungi (*Eumycota*) because there is one specie of the *Oomycota* class (light blue). Some findings support the hypothesis of certain mycologists that an extensive horizontal gene transfer had occurred between phytopathogenic ascomycetes and oomycetes early in evolution[11]. Despite all the progress in understanding the mechanism and all the data collected about UPOs, the natural function of these enzymes in fungal organisms is not clear yet. Obviously, the advantage of having a higher catalytic versatility may suggest an involvement in all kinds of detoxification reactions and other functions such as involvement in lignin and humus degradation or the modification of biosynthetic pathways[15,16]. Even more, yeasts such as *Saccharomyces spp.* don't have UPO genes. Also, neither plants (including green algae) nor animals have these genes[11].

More detailed analysis of UPO-sequences data revealed that there are two large groups of these enzymes: short and long peroxygenases[15]. Short UPO sequences (group I in *Figure 3*) are found in all fungal phyla and have a lower molecular weight than long UPOs (group II). The latter are present only in basidiomycetes and ascomycetes. There are also other differences between these groups in the active sites. For example, in group I, a conserved histidine acts as the charge stabilizer and in the group II (*Aae*UPO type), there's an arginine in this position.

UPOs are apparently organized in gene clusters (multigene families) in fungal organisms. Transcriptome studies on *Agrocybe aegerita* have indicated the presence of at least 16 UPO sequences probably including several gene variants[17]. Among the sequences there are both group II and a few group I UPOs.

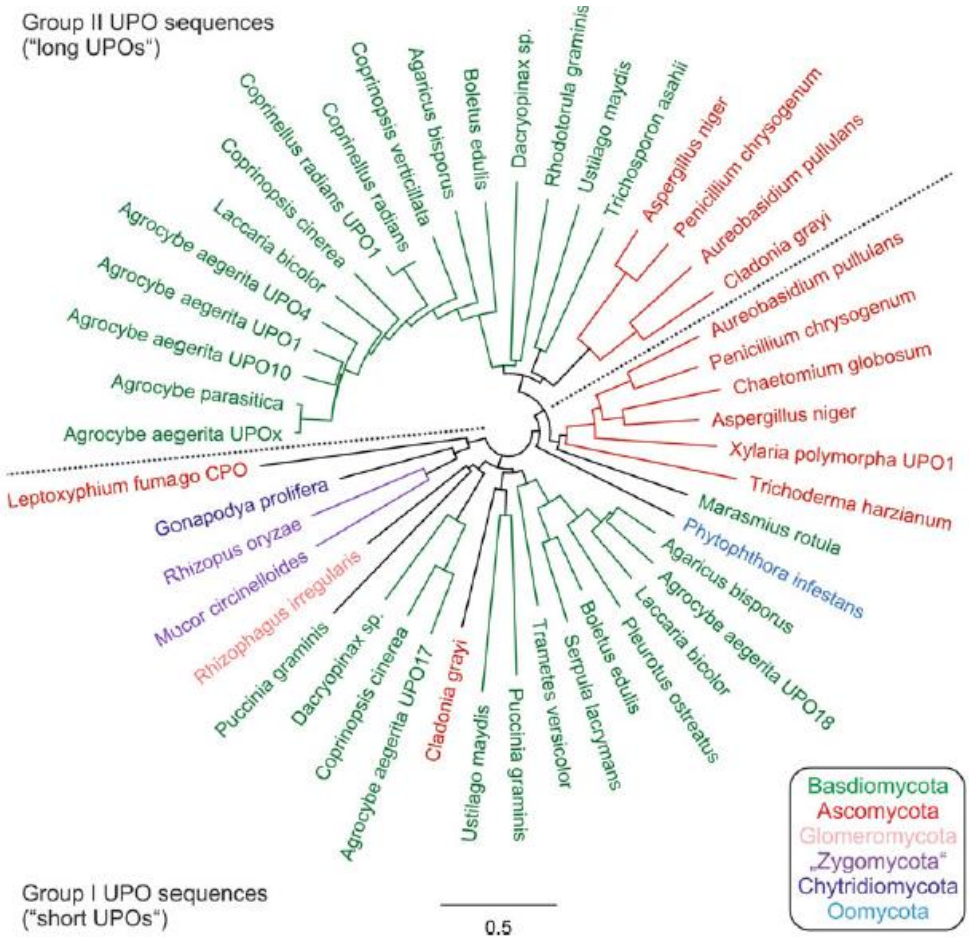


*Figure 3. Phylogenetic tree of UPO sequences. **Green** → Basidiomycota, **red** →Ascomycota, **light blue** → Oomycota, **purple** → Zygomycoya, **dark blue** → Chytridiomycota and **pink** → Glomeromycota. Dotted lines separate UPO groups I and II. Extracted from Hofrichter, M., et al. (2015) Fungal Unspecific Peroxygenases: Heme-Thiolate Proteins That Combine Peroxidase and Cytochrome P450 Properties. In Hrycay EG and Bandiera SM, editors, Monooxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome P450 (341-361). Springer International Publishing Switzerland.*

2.1.2. Structural characteristics of the *Aae*UPO

The UPO protein contains ten α-helices linked by long loops and five very short β-sheets. According to SCOP2, it belongs to the all α class of common heme-containing proteins. It has one disulfide bridge between Cys278 and Cys329 that stabilizes the C-terminal region after the last α-helix. Also, there are seven potential *N*-glycosylation sites of the high-mannose type. Its catalytic pocket is carved into the center of the structure through a long cone-shaped channel (17Å long) formed by hydrophobic/aromatic amino acids and surrounded by loops. The heme access channel is upholstered by 9 aromatic residues (eight Phe residues and one Tyr residue), of which Phe69, Phe121 and Phe199 form a triad that orients the substrate towards the heme, whereas Phe76 and Phe191 delimit the entrance to the channel[2,16].

The enzyme includes at the end of this channel, the protoporphyrin IX (heme group) with the iron hexa-coordinated, with its fifth position associated with the proximal Cys36 (axial) ligand and the sixth position (distal ligand) with a water molecule. It also has a structural magnesium ion that stabilizes the porphyrin ring system. It is coordinated by a glutamic acid (Glu122) and a serine (Ser126)[11,14,16].

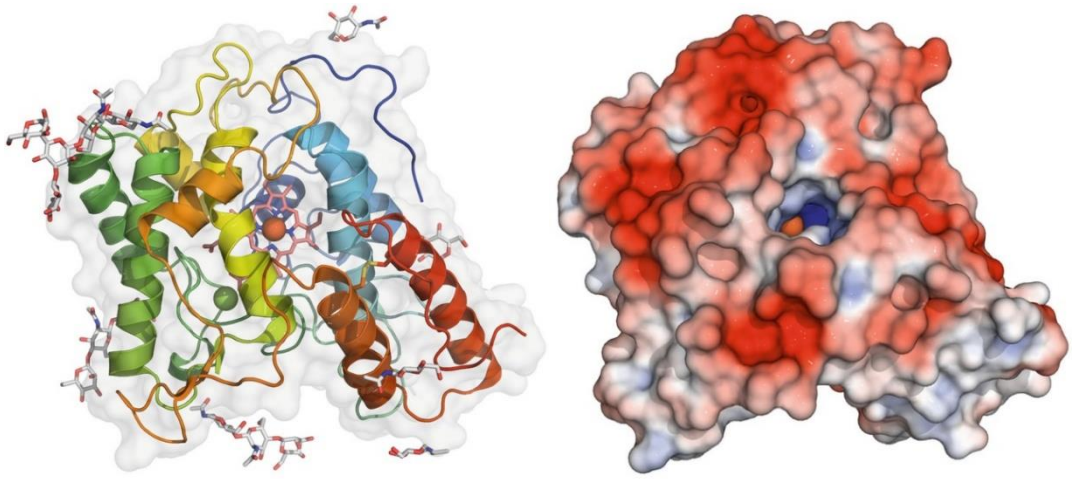


Figure 4. Structural traits of AaeUPO. Left: Ribbon diagram of the molecular structure of AaeUPO; α-helices (rainbow colored); heme iron (brown ball), magnesium (green ball). Right: solvent access surface of AaeUPO (colors represent electrostatic potentials: blue-positive, red-negative. Extracted from Hofrichter, M., Kellner, H., et al. (2020) Fungal Peroxygenases: A phylogenetically old superfamily of heme enzymes with promiscuity for oxygen transfer reactions. In Nevalainen, H., editor, Grand Challenges in Fungal Biotechnology (369-403). Springer Nature Switzerland AG.

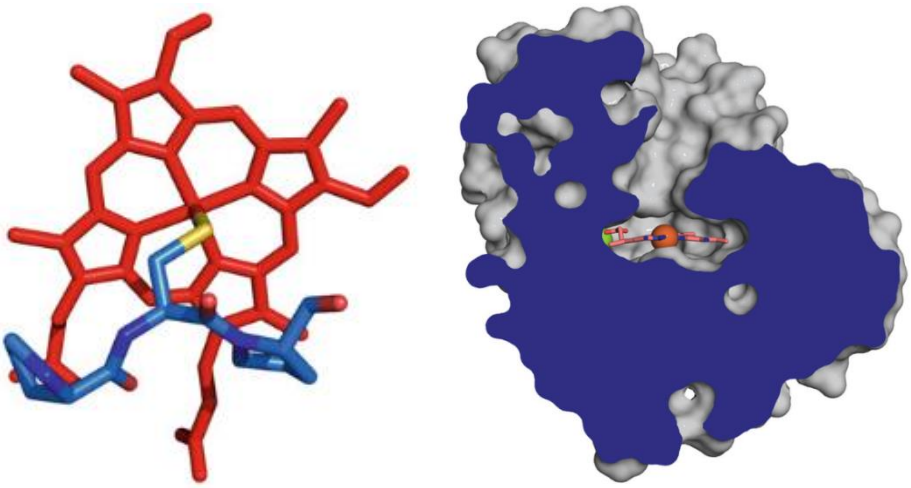


Figure 5. Left: Spatial arrangement of the PCP motif. Right: Tangential cross section through AaeUPO. Two proline residues (Pro35 and Pro37 in AaeUPO) expose the cysteine (Cys36) in a way that it can perfectly ligate the heme iron. (Left) Extracted from Hofrichter, M., et al. (2015) Fungal Unspecific Peroxygenases: Heme-Thiolate Proteins That Combine Peroxidase and Cytochrome P450 Properties. In Hrycay EG and Bandiera SM, editors, Monooxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome P450 (341-361). Springer International Publishing Switzerland. (Right) Extracted from Hofrichter, M., et al. (2020) Fungal Peroxygenases: A phylogenetically old superfamily of heme enzymes with promiscuity for oxygen transfer reactions. In Nevalainen, H., editor, Grand Challenges in Fungal Biotechnology (369-403). Springer Nature Switzerland AG.

The PCP motif that exposes a cysteine as the proximal ligand to the heme is highly conserved in most UPOs along with a particular glutamic acid residue (Glu196). The Glu196-Arg189 pair forms the distinctive acid-base pair required for the binding and the heterolytic cleavage of peroxide, which leads to enzyme activation via consecutive formation of the catalytic intermediates compound 0 (ferric-peroxo complex) and compound I (ferryl-oxo complex)[18]. The glutamic acid is deprotonated because it abstracts a proton from the iron-bound peroxide to form compound 0 and the charge stabilizer is Arg189. Compound I is the reactive key intermediate that can follow two different pathways depending on the substrate. On the one hand, it can follow a “two-

electron oxidation" of one substrate molecule along with oxygen atom transfer with peroxide as the oxygen source (peroxygenative activity). On the other hand, it can follow two "one-electron oxidations" of two substrate molecules resulting in the formation of the corresponding substrate radicals (peroxidative activity)[11].

2.1.3. AaeUPO's catalyzed reactions and reaction mechanisms

As said before, UPO shows both peroxidative and, more remarkably, peroxygenative activity. The latest one brings the promiscuity of UPOs in a wide portfolio of transformations that includes aromatic, alkylic, (cyclo)aliphatic, and heterocyclic hydroxylations, aromatic and aliphatic olefin epoxidations, ether cleavages (O-dealkylations), N-dealkylations, sulfoxidations, N-oxidations, deacylations (C−C bond cleavages), and halide oxidations/halogenations[11,19]. Also, more than 350 compounds have been demonstrated to serve as UPO substrates[19]. In *Figure 6* we can see the classification of the oxygenation and oxidation reactions.

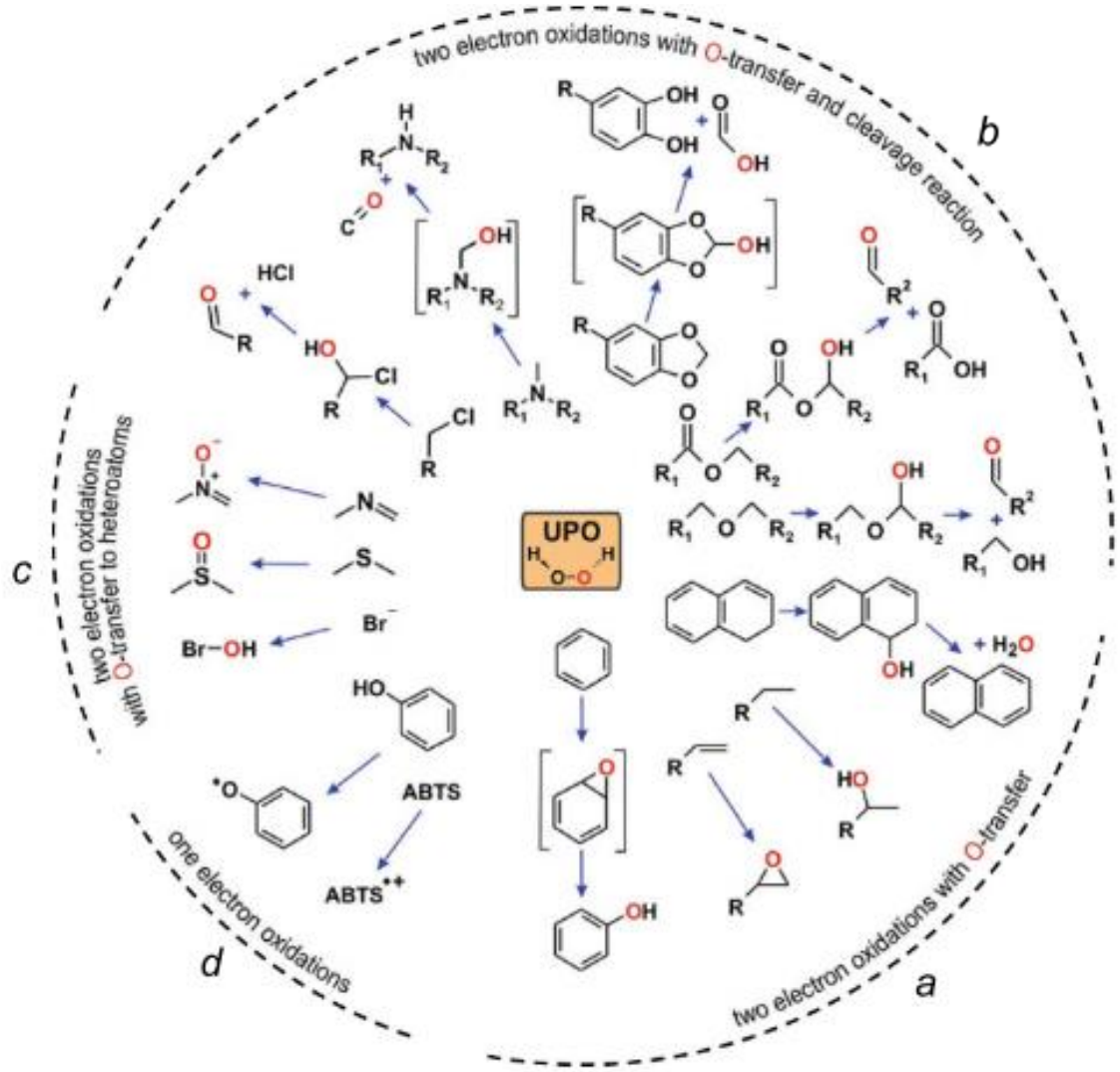


Figure 6. Summarizing overview of UPO-catalyzed reactions. UPO oxidation and oxyfunctionalisation at expenses of H2O2 include a) Two-electron oxidations with O-transfer; b) Two-electron oxidations with O-transfer and cleavage reaction; c) Two-electron oxidations with O-transfer to heteroatoms (S or N); d) One-electron oxidations. Extracted from Martinez, A., Ruiz-Dueñas, FJ., et al. Oxidoreductases on their way to industrial biotransformations. Biotechnology Advances 35, 815–831 (2017).

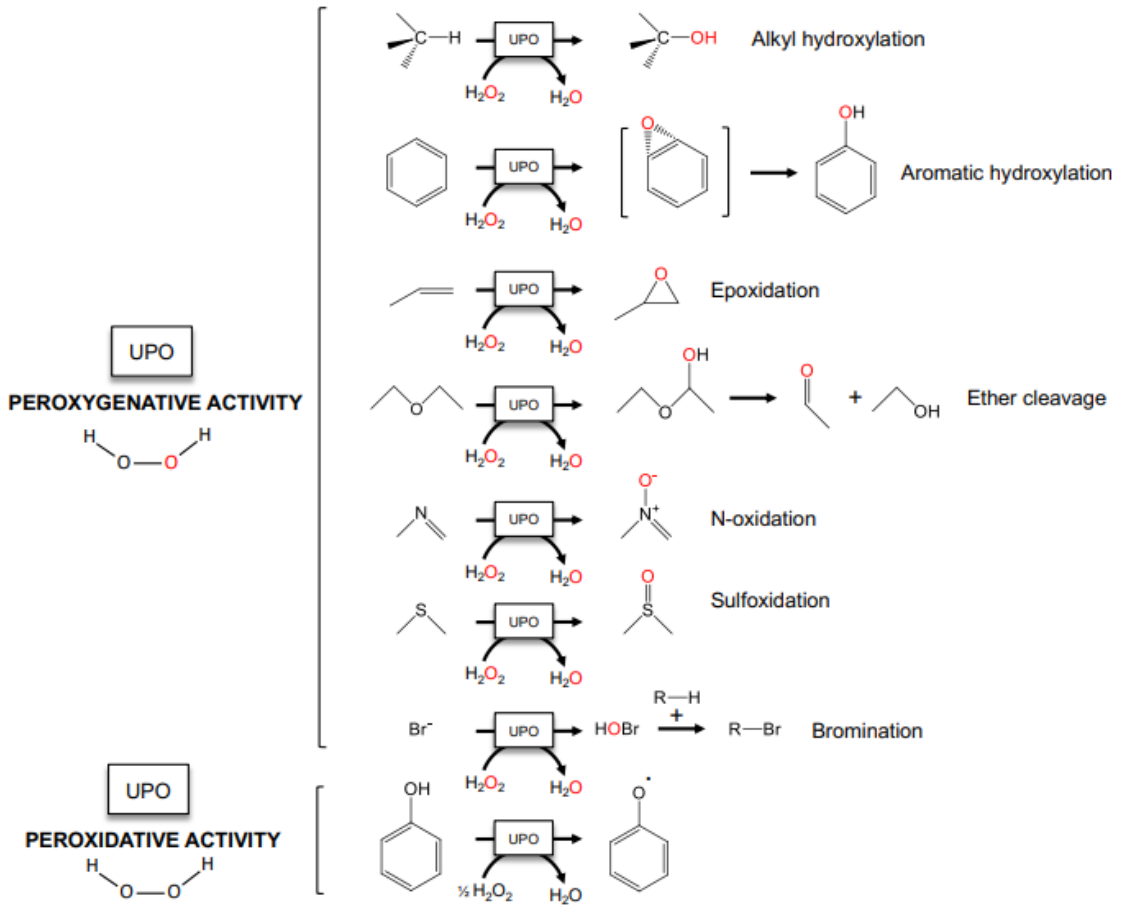


Figure 7. Representation of peroxygenative and peroxidative activities performed by UPOs. Extracted from Ramirez-Escudero, M., Molina-Espeja, P., et al. Structural Insights into the Substrate Promiscuity of a Laboratory-Evolved Peroxygenase. ACS Chem. Biol. 13, 3259−3268 (2018).

As said, compounds I and II are the key reactive intermediates in the proposed reaction cycle of UPOs. Resting state UPO (top middle structure in *Figure 9*) contains a ferric heme that has a loose bond water as the sixth (distal) ligand. In the first step of the **peroxygenative path** (left in *Figure 9*), the hydrogen peroxide (H_2O_2) replaces the water to form a short-lived pre-compound 0 (C0) (middle structure in *Figure* 9). Later, this C0 is deprotonated via a conserved glutamic acid (Glu196-COO^-) to form a C0 negatively charged a ferric peroxo-complex, not shown in *Figure 9*) [heme-Fe^{3+}-OO^-]. This C0 is heterolytically cleaved under electron re-arrangement to give the first metastable key intermediate: compound I (CI) an oxo-ferryl cation radical complex [$^{+\bullet}$heme-Fe^{4+}=O]. CI is a strong oxidant that will abstract a hydrogen atom from the substrate, yielding a protonated compound II (CII), a ferryl hydroxide complex [heme-Fe^{4+}-OH] and the substrate as a radical that stays in the active site close to the heme. This short-lived radical rapidly recombines to the corresponding hydroxylated product and the ferric enzyme (left side in *Figure 9*). While the product is released, a water molecule is coordinated to the ferric heme iron to form the initial state to be able to start the cycle again[20]. The cycle of the peroxygenase pathway varies when there is an epoxidation reaction. A modified compound II is proposed to bind the substrate as a radical via the ferryl oxygen (ferryl alkoxy radical complex). In this case, no H-abstraction would take place[11,13,18,19,21].

In the **peroxidase cycle** (right in *Figure 9*), starts also from compound I. Despite this, both compounds I and II abstract one electron each from the hydroxyl of two substrate molecules (in the example of the *Figure 9,* from the phenol). This results in the formation of two substrate radicals, that will be released and will undergo coupling and/or disproportionation reactions[20].

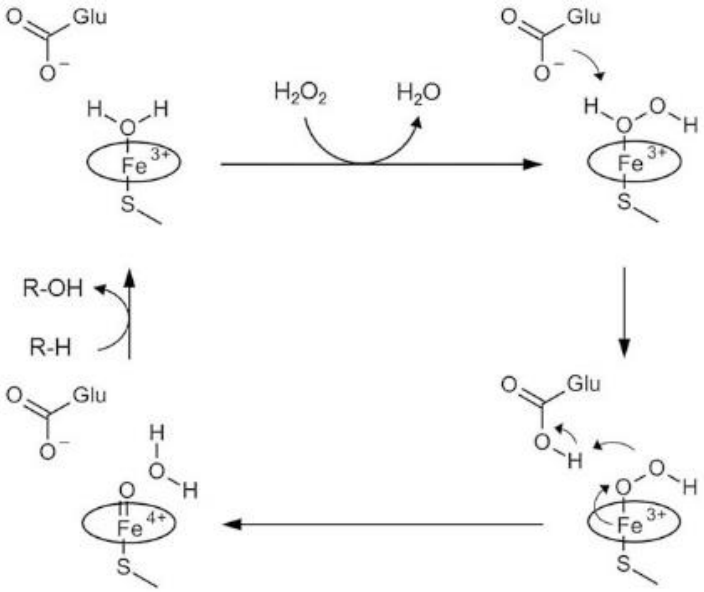


Figure 8. Formation of AaeUPO compounds 0 and I and role of deprotonated Glu196 that facilitates heterolytic cleavage and compound 0 formation. Extracted from Hofrichter, M., Kellner, H., et al. (2015) Fungal Unspecific Peroxygenases: Heme-Thiolate Proteins That Combine Peroxidase and Cytochrome P450 Properties. In Hrycay EG and Bandiera SM, editors, Monooxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome P450 (341-361). Springer International Publishing Switzerland.

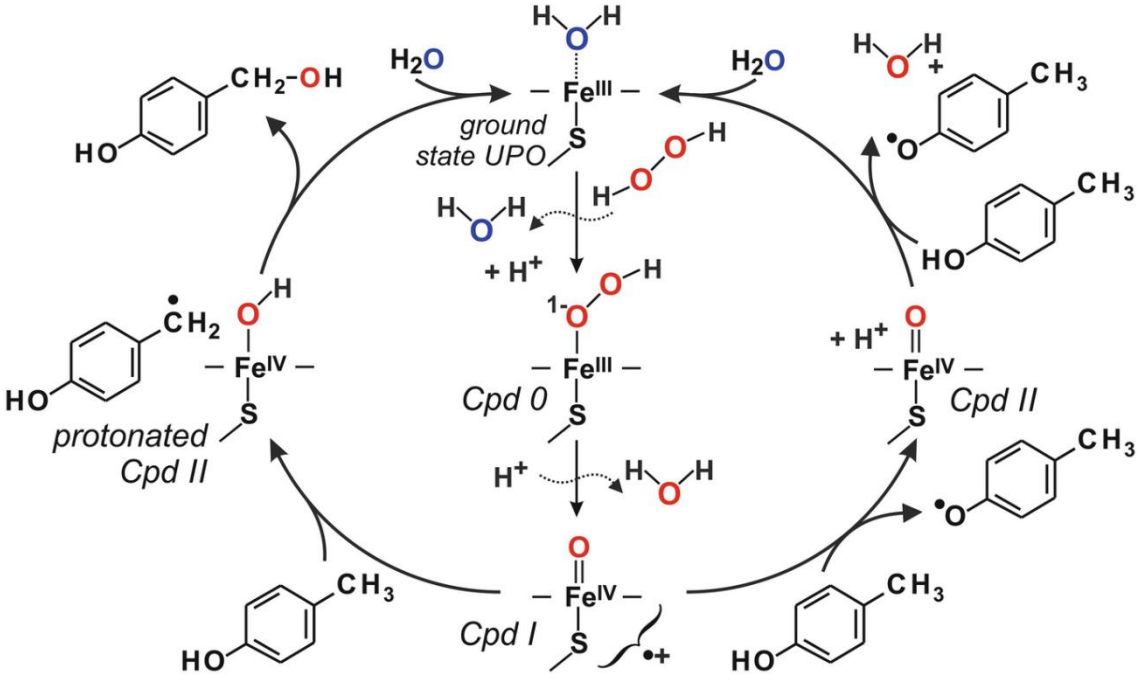


Figure 9. Proposed reaction cycle of UPOs with two routes: peroxygenase (left) and peroxidase pathway (right). Extracted from Hofrichter, M., Kellner, H., et al. (2020) Fungal Peroxygenases: A phylogenetically old superfamily of heme enzymes with promiscuity for oxygen transfer reactions. In Nevalainen, H., editor, Grand Challenges in Fungal Biotechnology (369-403). Springer Nature Switzerland AG.

2.2. PaDa-I: Directed evolution of the UPO from *Agrocybe aegerita*

The PaDa-I mutant is the outcome of five rounds of directed evolution from the wildtype enzyme of *Aae*UPO to enhance its activity and its heterologous functional expression in *Saccharomyces cerevisiae.* It has nine mutations (F12Y-A14V-R15G-A21D-V57A-L67F-V75I-I248V-F311L) where the underlined ones lie on the signal peptide. This evolved recombinant peroxygenase is more strongly glycosylated (33%) than the native *Aae*UPO (22%) when is expressed in some systems such as in *Pichia pastoris.* However, the kinetic parameters of PaDa-I were measured before and after a deglycosylation and it was found that the deglycosylated enzyme conserved its activity on both peroxidative and peroxygenative substrates[22–24].

The structure of PaDa-I is formed by the polypeptide chain Glu1-Arg327 (including the mutations previously mentioned on the backbone). It has an overall structural arrangement like the one described for the native *Aae*UPO.

As in the native *Aae*UPO, the C-terminus of the evolved PaDa-I is stabilized by the Cys278-Cys319 disulfide bridge while the N-terminus is markedly different. The mutant maintains the complete N-terminal region while the wildtype enzyme experiences proteolysis at Gly3-Leu4. This lack of proteolysis at the N-terminus of the recombinant enzyme could be due to the introduction of the V57A and V75I mutations. These substitutions are in the inner part of two helices that are near the N-terminal Pro5-Pro6 segment, and they produce a shift of 1Å at the Pro5 that could modify the hydrophobic contacts, possibly frustrating the processing[2].

Among other mutations, the L67F mutation is buried in a hydrophobic pocket defined by Phe204, Phe222 and Phe232 near the heme group, making hydrophobic contact with one of its methylene groups. Phe67 retains the same pattern of interactions with the neighboring aromatic residues, producing no significant evidence of structural changes, but it cannot be denied electronic effects in the protoporphyrin ring[2,23].

The I248V and F311L mutations are placed in positions with the side chains oriented towards the heme channel, shaping the active site and affecting the ligand binding. Ile248 is in a buried hydrophobic pocked far from the heme and its change with Val didn't result in structural modifications. The opposite happens with the F311L mutation because it resulted in the most significant change in the structure of PaDa-I, along with the prior mentioned (complete N-terminus). Phe311 is 3,6Å from Phe76, establishing a hydrophobic contact. There are two residues in the wildtype enzyme, Phe76 and Phe191, that form a pair of emerging aromatic residues at the entrance of the heme channel. This pair guide the substrate to the active site. The change of a bulkier residue such as Phe to a smaller Leu at position 311 shifts the Phe76 sidechain preserving the interaction with Leu311 while weakening the contact with Phe191. This change leaves Phe191 more exposed and reorientated to stablish an interaction with Phe274 at 3,7Å. The overall consequence is a larger entrance to the heme channel[2,23].

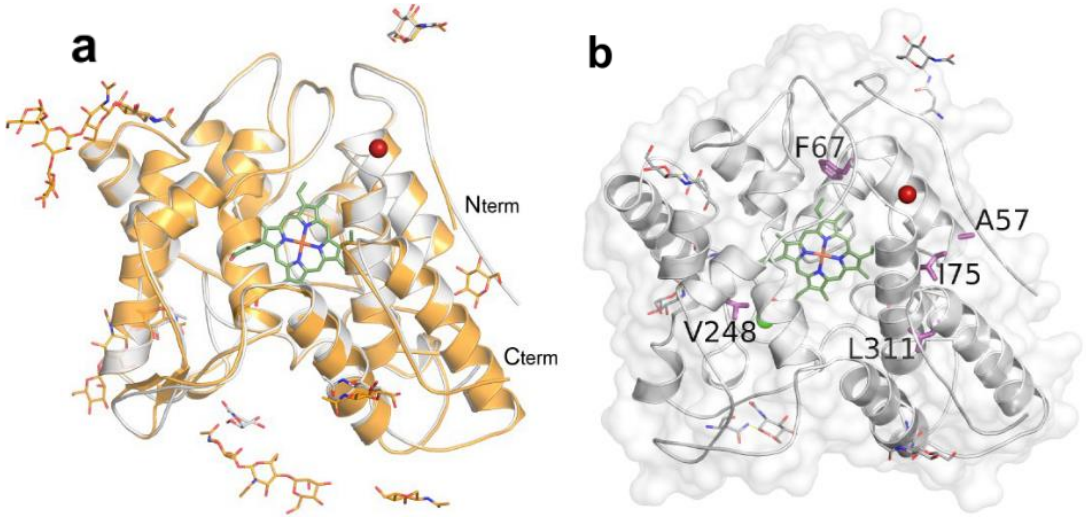


Figure 10. General view of the evolved peroxygenase structure. (a) The cartoon represents the overall structure of native (light orange) vs PaDa-I (white), showing the magnesium ion (Mg^{2+}, green sphere), the heme group (dark green sticks), the chloride ion (red sphere), and the different glycan chains as sticks. (b) The residues mutated in the evolved peroxygenase highlighted as purple sticks. Protein Data Bank entry 5OXU. Extracted from Ramirez-Escudero, M., Molina-Espeja, P., et al. Structural Insights into the Substrate Promiscuity of a Laboratory-Evolved Peroxygenase. ACS Chem. Biol. 13, 3259−3268 (2018).

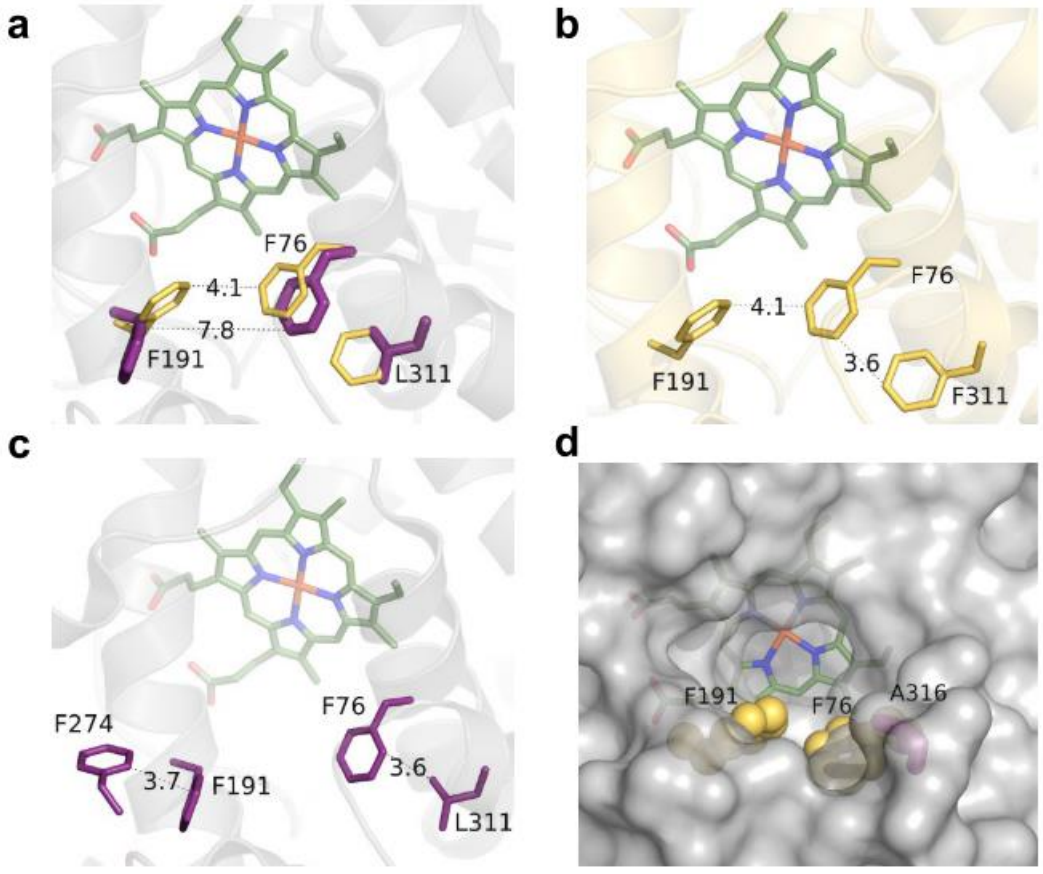


Figure 11. Active site differences between the mutant and the wildtype peroxygenase. (a) The F311L change (wildtype AaeUPO, yellow; PaDa-I, purple) produces a shift in Phe76 that affects the Phe191 sidechain, expanding the entrance to the heme group (green). (b) wild type AaeUPO or (c) recombinant PaDa-I crystal structures. (d) The surface representation illustrates the channel. The residues in yellow (wildtype) show the narrowed entrance. Protein Data Bank (PDB) entry 5OXU; AaeUPO, PDB entry 2YOR. Extracted from Ramirez-Escudero, M., Molina-Espeja, P., et al. Structural Insights into the Substrate Promiscuity of a Laboratory-Evolved Peroxygenase. ACS Chem. Biol. 13, 3259−3268 (2018).

2.3. PELE software

PELE *(Protein Energy Landscape Exploration)* (https://pele.bsc.es), is a method developed to perform protein energy landscape explorations combining a Monte Carlo (MC) stochastic approach with protein structure prediction algorithms[25]. This software is capable of accurately reproducing long time scale processes of protein or protein-ligand conformations in few hours of CPU.

The application of MC techniques in biomolecule sampling is not a straightforward procedure because of the difficulties in applying global sampling moves. Even more, the protein-ligand energy landscape is full of local minima due to the huge number of degrees of freedom. On the one hand, ligands contain rotatable bonds and during the interactions with other molecules they can adopt different conformations. On the other, protein backbones and sidechains are flexible and introduce many different states. For all these reasons, even with the binding site identified previously, the number of possible configurations to consider is enormous. In this context, PELE was designed to enhance MC sampling in such difficult cases[25,26]. Its key contribution resides in using protein structure prediction techniques coupled to random tests, enhancing the MC moves toward important sampling regions. First developed to sample and study the conformational space between proteins and ligand, PELE has been applied to study several ligand migrations, induced fit docking and protein dynamics with less computational cost than molecular dynamics (MD).

Each PELE MC step is composed by three main parts: localized perturbation, side chain sampling and minimization.

On the **localized perturbation,** after an energy calculation for the initial structure, the procedure begins with the generation of a perturbation in the system. In studies of ligand diffusion or ligand induce fit, the perturbation starts with a random translation and rotation of the ligand. The perturbation can include the backbone of the protein (or the backbone close to the ligand) and later a minimization is performed (in all the atoms) where the α-carbons are driven to a new position resulting from a small displacement in a low frequency anisotropic normal mode (ANM). This procedure aims to describe the global motion of the protein taking the initial conformation from a local minimum generating another configuration. On the **side chain sampling**, the algorithm places all side chains local to the ligand with a rotamer library side chain optimization. The side chain algorithm uses steric filtering and a clustering method to reduce the number of rotamers to be minimized (side chains minimization only). The last step is the **minimization,** and it involves the minimization of a user-defined region (typically the whole protein is included) with a Truncated Newton minimizer.

These three steps compose a move that can be accepted (defining a new minimum) or rejected based on a Metropolis criterion to approximate a Boltzmann distribution of the PELE steps in terms of energy. Whenever any trajectory is significantly further along a given reaction coordinate than any of the other trajectories, the route trajectory is abandoned and restarted from the position of the leading trajectory[25]. This allows an

efficient sampling of the configurational space towards one defined objective (entrance/escape of the ligand, optimization of the protein-ligand binding energy, etc).

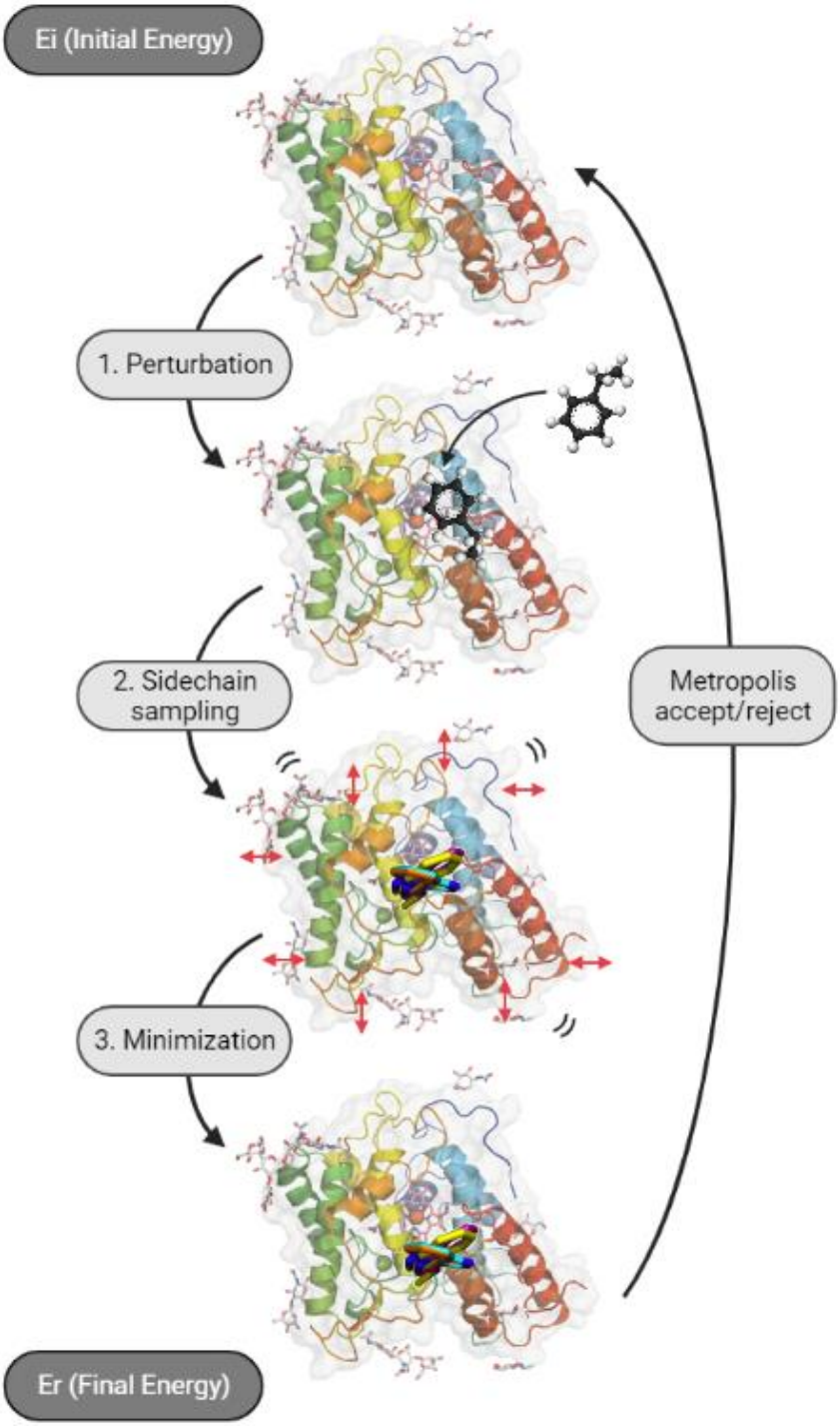


Figure 12. PELE scheme.

For the present work, the adaptive version of the algorithm has been applied. AdaptivePELE is a Python module that performs an enhanced sampling of the molecular simulation built around PELE[27,28]. It works running short simulations assessing the exploration with a clustering and spawning new trajectories in "interesting" regions. This "version" is composed of three main steps: sampling, clustering, and spawning, which

are run iteratively. In the **sampling**, a series of independent trajectories are run. These trajectories are generated with the classical PELE approach. We use rounds (named epochs) of a determinate number of simulations (trajectories) of a determinate length, each one running on a computing core. Then the **clustering** phase cluster all the conformations generated in all previous epochs. Typically, ligand RMSD is used as the metric for clustering. Each cluster has a central conformation and a similarity RMSD threshold (because of this, a structure belongs to a cluster if its RMSD with the central conformation is smaller than the threshold). When a structure does not belong to any cluster, a new one is created, defining a new center. Finally, the **spawning** phase chooses the seeds to be used for the next iteration (next epoch). The simulations are stopped, and an adaptive spawning is performed with new initial structures for the next iteration. Of course, in the selection strategy, the sampling can be biased to interesting areas known.

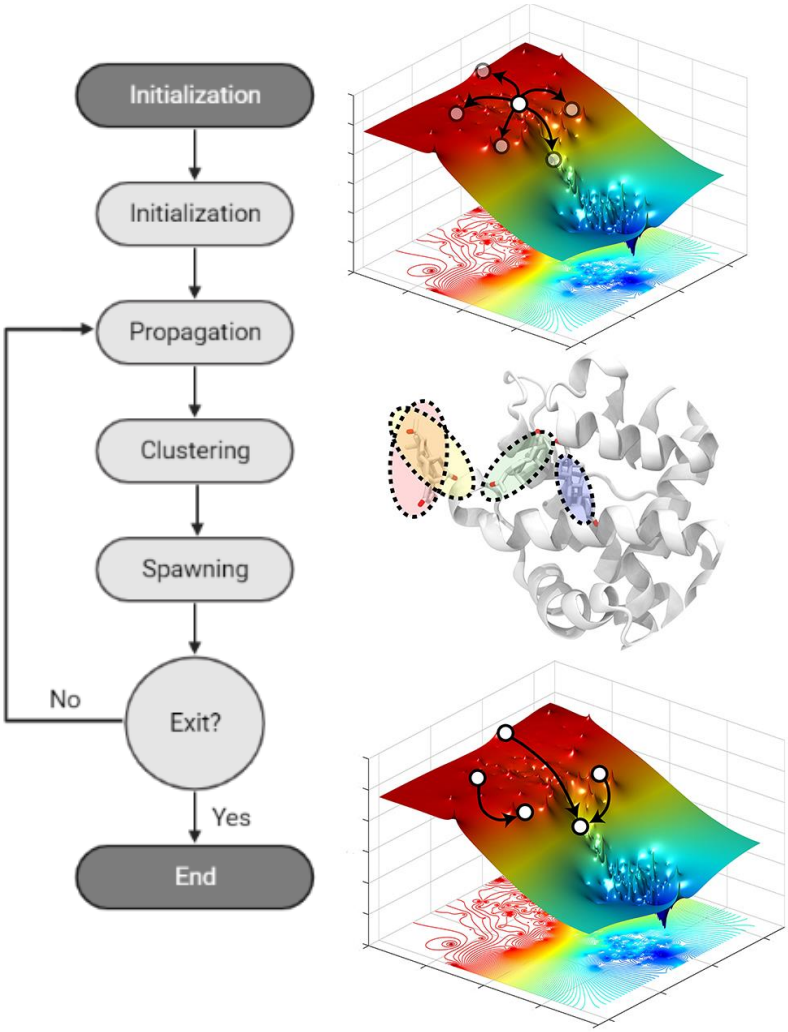


Figure 13. Flow diagram that summarizes the algorithm. Adapted from AdaptivePELE's documentation. (https://adaptivepele.github.io/AdaptivePELE/) [15 August 2021].

PELE algorithm has been applied to study ligand diffusion, induced fit docking, protein local motion and absolute binding free energy estimation, among several others[29–33]. The use of the PELE software has a strong background of presented consistent experimental and theoretical data. PELE simulations properly inform about substrate

access and positioning at the enzyme active site including distances and angles between the redox centers as stated in previous studies for UPO reactions[29,31,32,34]. Even more, in several studies, using the analysis of the substrate diffusion, it is possible to estimate the binding free energies. Furthermore, the predicted values show a noticeable correlation with the turnover numbers determined experimentally.

2.4. UPOs' applications. Its current state-of-the-art

One of the obstacles that still stands in the way of industrial UPO applications is their limited availability and the outcome of the specific reactions desired. Successful attempts in both directions have been made. These enzymes are highly promising biocatalysts for a variety of aromatic and aliphatic oxyfunctionalizations of industrial interest and, because of this, its homologous and heterologous production has been extensively studied. In some UPOs, proper levels of expression have been achieved. In combination with the whole set of modern protein engineering techniques, it will be possible to construct and express tailor-made peroxygenases for specific synthetic problems in the near future[20].

Talking about conducting UPOs to catalyze reactions of interest, in recent studies, UPOs were submitted to directed evolution to create variants or mutants. One case is the one in which their peroxygenase activity was enhanced while its one-electron oxidation (peroxidative activity) reduced[2,22,35]. In addition, other variants as PaDa-I have been engineered. This are the cases of JaWa, SoLo, Qui-gon and WinDu. Some variants have been used in oxyfunctionalizations of biotechnological interest like dye precursors, drug ingredients, metabolites, and renewable chemical building blocks. Here we have a few more examples:

> An amount of 40.000 tons/year of 1-naphtol are used in the chemical industry in the production of herbicides, insecticides, pharmaceuticals, and dye precursors. This chemical can be obtained by naphthalene hydroxylation (it's an epoxidation followed by a spontaneous re-aromatization) using a UPO variant evolved for this purpose[36]. Other epoxidations of non-cyclic alkenes and terpenes of interest has also been achieved using different both wildtype and recombinant UPOs[10,19,37].
>
> 25-hydroxyvitamin D3 is a product of interest in feeding chickens and other farm animals to reduce skeletal problems caused by rapid growth and reduced mobility. This component can be optimally produced using the recombinant *Coprinopsis cinerea* UPO (*Cci*UPO) from Novozymes[38]. Other similar reactions had been applied on a variety of steroidal substrates, mainly at their 25 position, producing molecules with antimicrobial properties, among others[39].
>
> The *N*-demethylation in the synthesis of drug metabolites of the bile acid reabsorption inhibitor SAR548304 (Sanofi) typically required a seven-step chemical reaction using palladium catalysis and laborious chromatographic purification with a slight yield. However, the same reaction can be accomplished with a higher yield using the *Marasmius rotula* UPO (*Mro*UPO)[40]. Also, in this

field, the same UPO is used for the enzymatic removal of corticoid sidechains, involving hydroxylation and C-C bond cleavage[41].

The oxyfunctionalisation of saturated hydrocarbons under mild conditions is a major challenge of modern chemistry because very few reagents can carry out the selective oxidation of alkanes. Because of this, enzymatic oxygenation of long inert *n*-alkanes at their unreactive terminal position has been studied and reported using the *Mro*UPO[42]. Also, stereoselective hydroxylation of hydrocarbons is also a challenging task for chemical catalysts. For example, the complete and stereoselective conversion of ethylbenzene into (R)-1-phenylethanol has been reported as a model reaction[3,8,38].

As seen, in most of the studied reactions, the oxidative enzymes were not suitable for industrial application as they are produced in nature, and their catalytic and/or operational properties need to be optimized by application of rational design and/or directed evolution tools. Also, this optimization could have additional economic and environmental benefits for the industry. Oxidoreductase engineering has benefited from computational simulations, where the target reaction to be achieved or the enzyme property to be improved was explored with *in silico* biophysical and biochemical tools. Because of this, the time dedicated to experimental engineering work can be significantly reduced.

Biophysical modelling typically included dynamic simulations of substrate diffusion to the enzyme active site, optimized docking and (if it can be a limiting step) analysis of product diffusion to the solvent region. For this purpose, the Protein Energy Landscape Exploration (PELE) software, an algorithm for modeling substrate and product migration in enzymes that is capable of accurately mapping long time scale processes in only few hours of CPU has been used[28,29,31,32].

Some examples of the use of PELE in the field of the UPOs can be named. It was used to explore the epoxidation of unsaturated fatty acids by *Aae*UPO, *Cro*UPO, and *Mro*UPO and some engineered variants[37]. Also, it was utilized to understand the yield and selectivity differences between different steroids in *Aae*UPO, *Mro*UPO, and *Cro*UPO variants[38]. Additionally, in the previous study mentioned about the 25-hydroxyvitamin D3 PELE was used to rationalize the results evaluating the diffusion of the D2 and D3 substrates[43]. In the same way, PELE was used to assess the peroxygenase activity of the mutations introduced in some *Aae*UPO variants in the synthesis of 1-naphtol[36]. Finally, as an example of the current line of work, the software was also used to explore the role of the phenylalanine residues in the modulation of the flexibility of the active site of the variant PaDa-I[23].

Next steps to explore and overcome are to continue searching for the optimization in the catalyzation of interesting substrates refining the existing mutants as PaDa-I. In the present work, this is the central focus. We aim to do a horizontal research by evaluating the binding energies, the regio-, and, if it is possible, the enantioselectivity of PaDa-I's variants in relation to five different substrates.

3. Objectives

Following the introduction, considering the relevancy and direction of the topic nowadays, current studies are centered in the development of very specific variants for concrete reactions. In our case, the study of PaDa-I's variants with the five substrates that will be stated will enlarge the data existing and it is going to set a starting point for future studies with more different substrates. Research in this line will prepare us for the design of new specific mutants for certain types of substrates that would be able to be expressed and produced in homologous and heterologous systems.

In this context, PELE act as a catapult because it permits to study the systems *in silico* without the need of wasting expensive resources and time. Even more, because of its background and experience in the field and in the specific systems (UPOs), it is presumed that the software PELE is a reliable tool.

With all of this, in this work we aim to analyze PaDa-I's mutants with PELE software in terms of binding energy, regioselectivity (in the substrates that can be carried out) and enantioselectivity. In a more schematic way, this would be:

- To test if PELE software can retrieve differences between the systems and substrates assessed.
- To evaluate PaDa-I's mutants against five substrates in terms of binding energy with a binding pose refinement using PELE software. We want to stablish which mutants show a greater value for every substrate.
- To assess PaDa-I's mutants' regioselectivity for ethylbenzene and 1,2,3,4-tetrahydronaphtalene substrates.
- To refine and construct a protocol to be able to estimate PaDa-I's mutants enantioselectivity for the five different substrates and applying it.

4. Materials and methods

4.1. System preparation for molecular modeling

The structure for the computational simulations of PaDa-I was provided by Miguel Alcalde's research group and isn't published on the PDB. Since this structure hadn't the oxygen placed on the heme group as compound I, this atom was taken from a published structure (PDB ID: 5OXU) from PaDa-I. All 30 mutants' structures were created from the structure of PaDa-I following the reference given and were prepared at a pH of 6 (same conditions used for the experimental analysis), using Schrödinger's Protein Preparation Wizard. Under these acidic conditions, histidines were double-protonated, except for His-82 and His-22 (ε-protonated), and His-118 and His-251 (δ-protonated), and all acidic residues were deprotonated. In all cases, structural water molecules (with 2+ hydrogen bonds with the protein) were included in the models. Glycosylations were removed because they didn't affect to the activity of the enzyme as stated previously[2,22,24]. The heme site was modeled as the two-electron-oxidized compound I, and its atomic parameters were obtained with a quantum mechanics/ molecular mechanics (QM/MM) minimization using Qsite from another work[37]. The force-field parameters of the heme CI in UPO were previously reported[37].

The five substrates' structures assessed (cis-β-methylstyrene, *trans*- β-methylstyrene, ethylbenzene, styrene and 1,2,3,4-tetrahydronapthalene) were obtained from PubChem (https://pubchem.ncbi.nlm.nih.gov/) and prepared with Schrodinger's Ligprep. They were optimized with the default protocol.

4.2. Docking procedure

Once the protein structures had been prepared and ligands optimized, a docking procedure with Schrodinger's Glide was performed. The ligands were initially placed at the entrance of the heme channel and the default settings, and the single precision SP option were used. The number of poses retrieved was 5 per procedure.

4.3. Adaptive PELE computational analysis

Once the docking procedure was performed, ligand pose refinement was performed with PELE. PELE is a heuristic Monte Carlo (MC) based algorithm that combines an initial random perturbation where the ligand is translated/rotated and the backbone is displaced using normal modes, with a relaxation second phase including protein structure prediction algorithms: side chain prediction and minimization. In this study the enhanced-sampling version, adaptive-PELE was used. A total of 4 epochs of 100 PELE steps were conducted in 96 processors, at the MareNostrum supercomputer (www.bsc.es/marenostrum). After each epoch, trajectories were clustered and new explorations were spawned on the most poorly explored regions to increase sampling efficiency.

The initial position for the 5 substrates was obtained from the Glide docking inside the cavity. All the mutants were considered to start from the same poses taking as reference the ones from PaDa-I. Ligands were already near the heme catalytic oxygen and were free to explore the active site pocket with a 6Å radius restrain. The simulation

parameters were chosen as “induced fit exhaustive” with the changes mentioned in the number of epochs and steps. Also, α-carbon constraints parameters were set to level 3 in which the whole backbone is constrained every 5 atoms with 5 kcal/mol. The results presented are based on 95 trajectories for every epoch (380 trajectories) in about 6 hours for every combination of mutant and ligand.

At each PELE MC step, the substrate was perturbed by applying a translation comprised between 0.5 and 1.25Å and a rotation angle between $2\pi/5$ and $\pi/10$ radians. Protein and substrate were described at the OPLS2005 force field level of theory.

One of the ways by that PELE results are analyzed, are by plots combining the binding energy and the RMSD of the ligand atom defined. The binding energy is computed as the internal energy difference between the complex and the free receptor and ligand.

5. Results

In the first part of the study, we analyzed the diffusion of five substrates: cis-β-methylstyrene (CMS), *trans*- β-methylstyrene (TMS), ethylbenzene (ETB), styrene (STR) and 1,2,3,4-tetrahydronapthalene (THN) with the thirty mutants given plus PaDa-I at the vicinity of the heme cofactor, using AdaptivePELE.

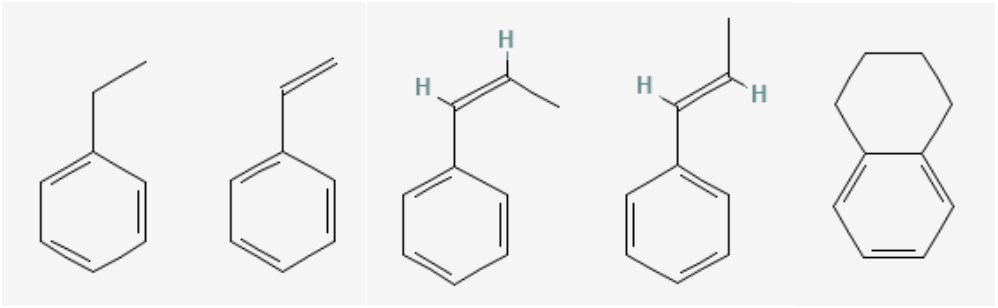


Figure 14. Substrates assessed. From left to right, ethylbenzene, styrene, cis-β-methylstyrene, trans-β-methylstyrene, 1,2,3,4-tetrahydronaphtalene.

The computational results consist of density and energetic plots of distances versus interaction energies. These distances are the ones between the oxo ligand in the simulated CI of the thirty-one structures and the catalytic carbon of the five substrates above mentioned. The specific distances computed were the ones seen in *Figure 15,* where for the substrates proposed to retrieve an epoxidation was calculated a new distance, the apothem of the triangle formed. This is the case for styrene, cis-β-methylstyrene, *trans*- β-methylstyrene and for the orto and para positions for ethylbenzene. For ethylbenzene, three distances were computed and for 1,2,3,4-tetrahydronapthalene four distances were computed.

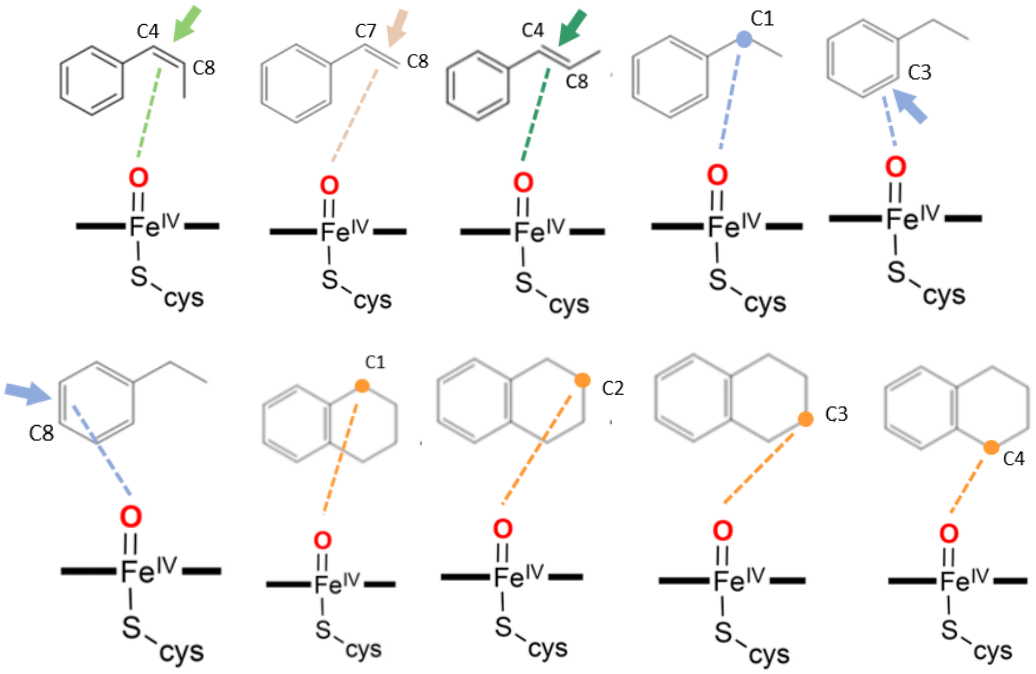


Figure 15. Distances computed for every substrate. In light green → cis-β-methylstyrene, in light brown → styrene, in dark green → trans-β-methylstyrene, in blue →ethylbenzene, in orange →1,2,3,4-tetrahydronaphtalene. The arrows point to the double bonds subject to be a future epoxidation and for the ones the distance is computed as the apothem.

In terms of raw results, we have a plot for every combination of substrate and mutant. In these plots, there are represented the binding energies (in kcal/mol) (x axis) and the distance (in Å) between the selected atoms (y axis) as can be seen in *Figure 16.* Also, in this example there is a red line to mark the 5Å distance because we considered it as the threshold of catalytical relevancy.

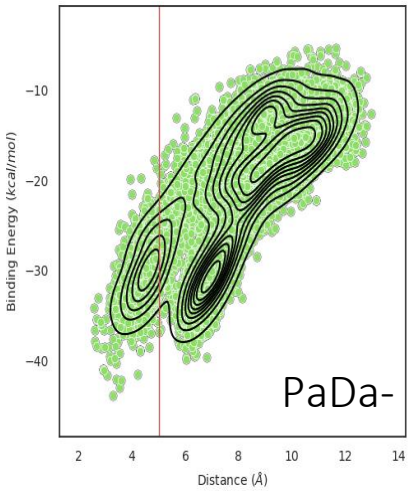


Figure 16. Plots obtained for every combination of substrate and mutant.

In analogy with any spontaneous process, protein-ligand binding occurs only when the change in Gibbs free energy of the system is negative. This binding energy is calculated as the energy of the complex (substrate-enzyme) minus the energies of the substrate and the enzyme separated. Because of this, a more negative value will denote a better binding or affinity. Also, considering the distance, the lower that it would get, the better the interaction because it would be a more catalytically relevant pose. Due to this, we are interested in the lower-left corner of the plot.

$$\Delta G = Gfinal - Ginitial$$

To be able to distinguish between mutants in terms of energy binding and distances, we analyzed the plots in thresholds (Section 5.1) and residence times (Section 5.2). The construction of the thresholds is explained in its section as well as the build of the residence times. In the case of the thresholds, a higher level of dots density in areas close to the lower-left corner will correspond to an interesting mutant. Also, a higher percentage value in the residence time would mean that the substrate has stayed in a catalytic position the percentage described, and thus, it would be a more interesting mutant.

5.1. Binding energies. Thresholds made of the plots

The plots were analyzed selecting three thresholds defined by energies and distances between the studied atoms. They are constructed as the percentage of the number of accepted steps (dots in the plot of *Figure* 17) that fall in each area considered divided by the total of accepted steps (and making it a percentage). The first threshold counts poses retrieved under -30kcal/mol and under 5Å, the second under -35kcal/mol and 4Å, and the third under -40kcal/mol and 3Å. In this way, the plots can be represented and compared more easily.

For ethylbenzene and 1,2,3,4-tetrahydronaphtalene there are more than one plot because of the different carbons assessed. For ethylbenzene we have C1, C3 (orto position) and C8 (para position). For 1,2,3,4-tetrahydronaphtalene we have two plots, one for C1 and 4 together because they are symmetric, and other for C2 and C3 as can be seen in *Figure 15.* Here, there are only shown results with higher values than PaDa-I (or Mutant 0).

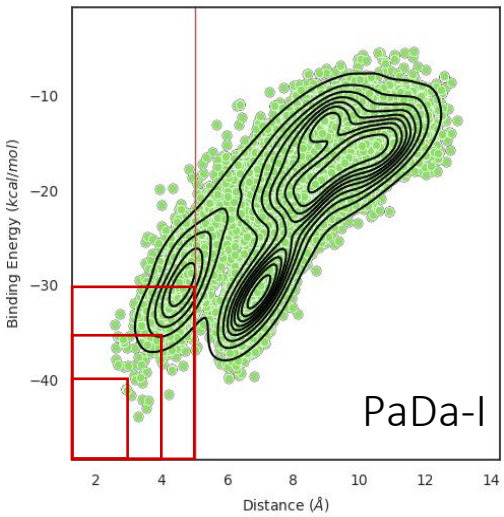


Figure 17. Representation of the areas (thresholds) used. This are a way to interpret the plots in an easier way. The first one includes poses between -30kcal/mol and 5Å of distance between the oxo and the C of interest. The second includes poses between -35kcal/mol and 4Å. The third one includes poses between -40kcal/mol and 3Å.

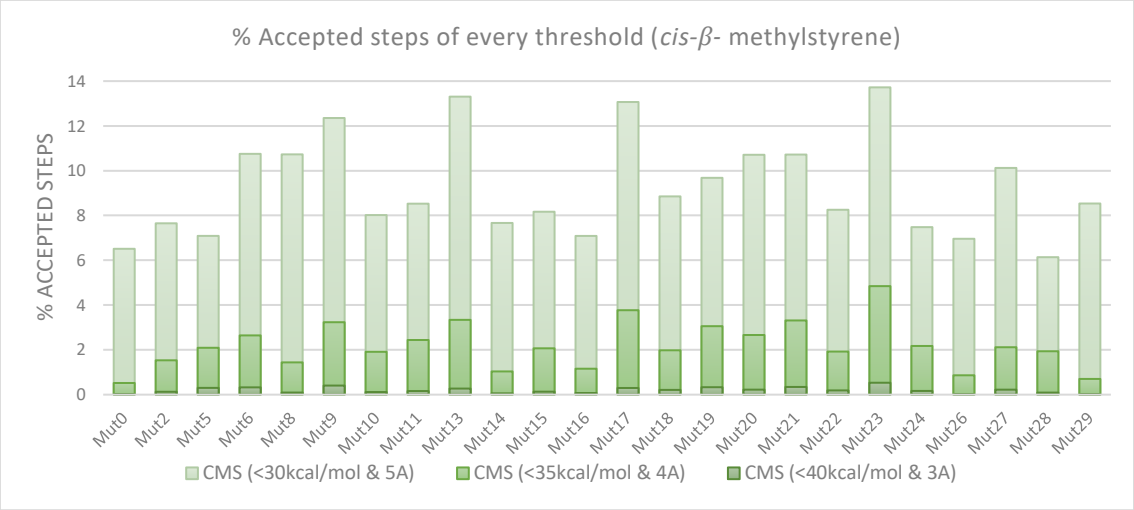


Figure 18. Thresholds of the accepted steps with cis-β-methylstyrene. There are only represented mutants with values higher than PaDa-I (Mut0).

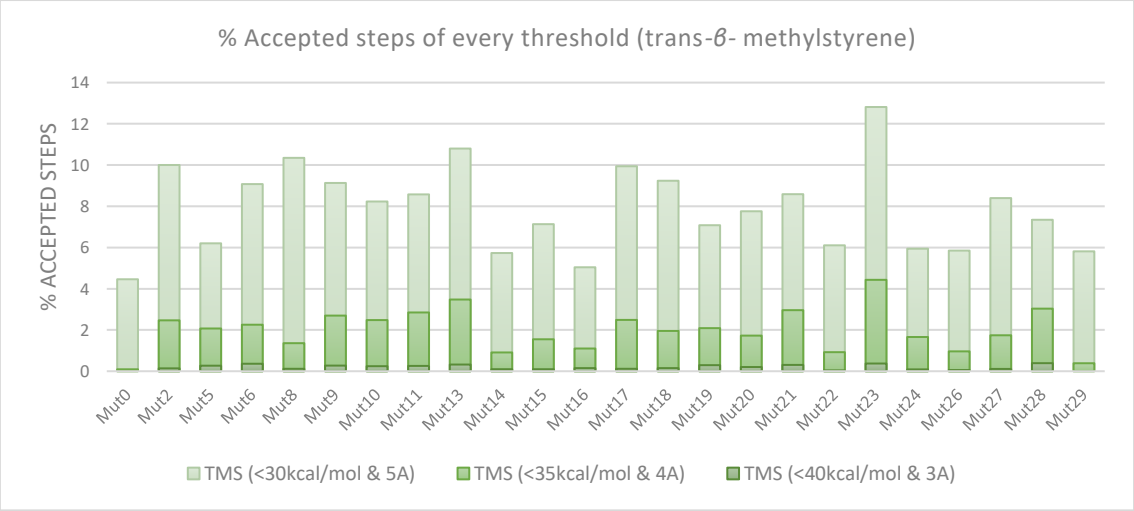


Figure 19. Thresholds of the accepted steps with trans-β-methylstyrene. There are only represented mutants with values higher than PaDa-I (Mut0).

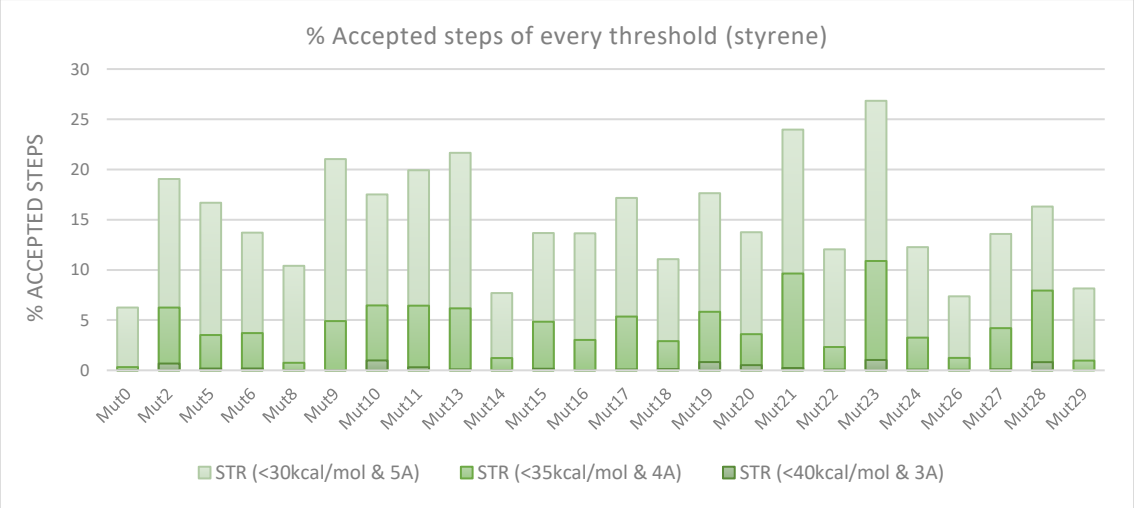


Figure 20. Thresholds of the accepted steps with styrene. There are only represented mutants with values higher than PaDa-I (Mut0).

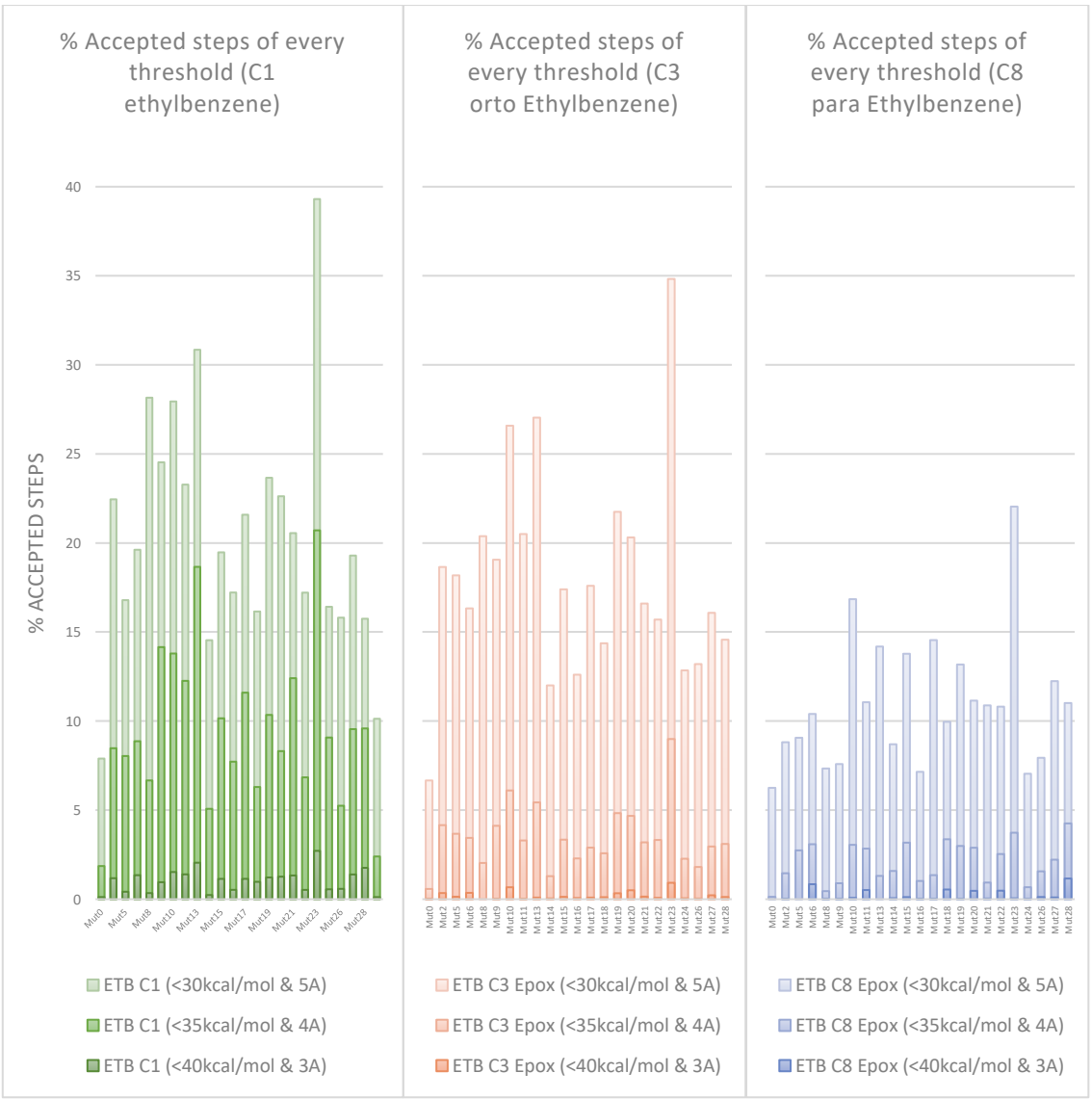
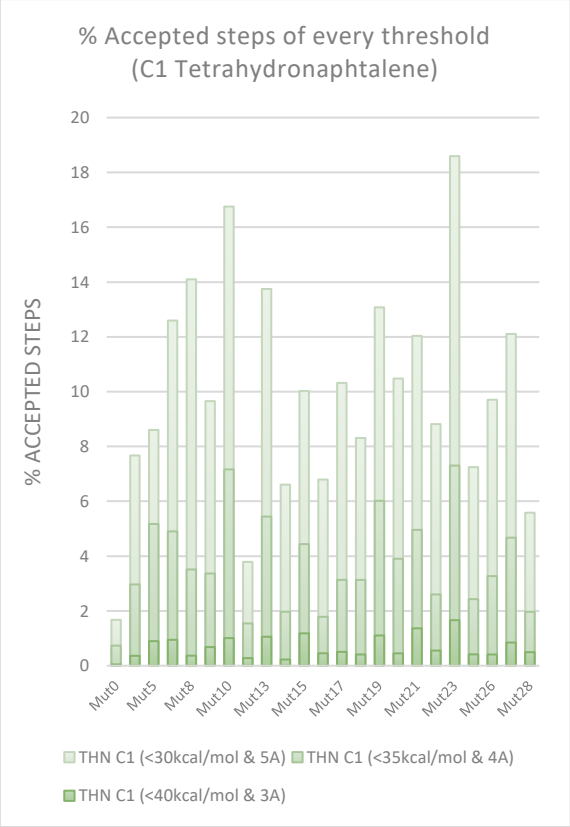


Figure 21, Figure 22 and Figure 23. Thresholds of the accepted steps with ethylbenzene (C1, C3 (orto) and C8 (para)). C1 in green, C3 in orange and C8 in blue. There are only represented mutants with values higher than PaDa-I (Mut0).



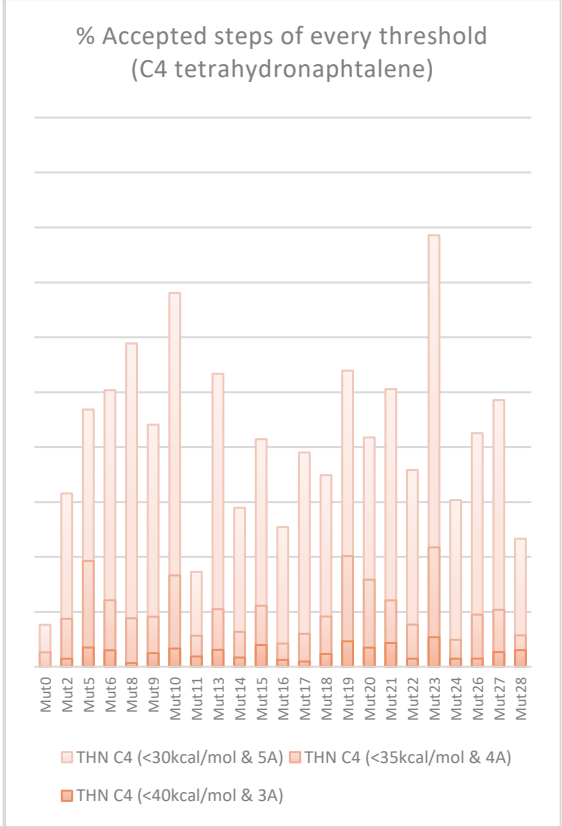
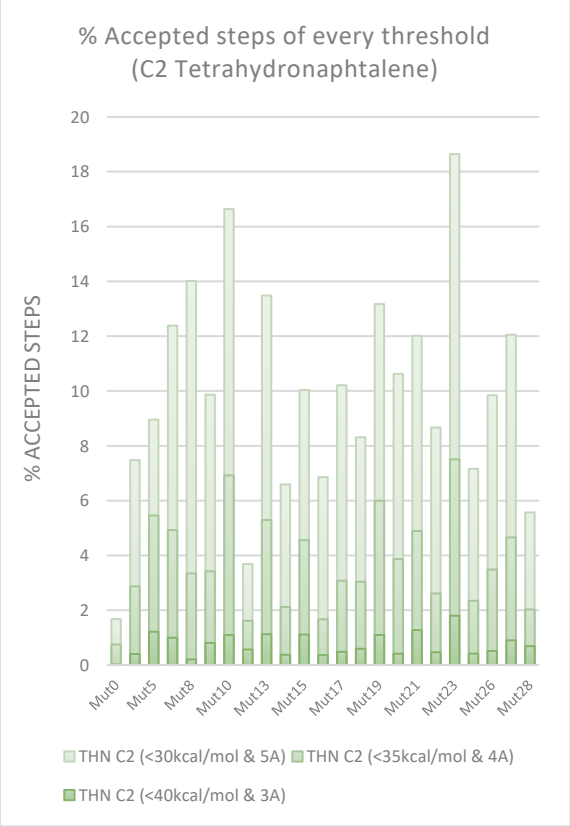


Figure 24 and Figure 25. Thresholds of the accepted steps with 1,2,3,4-tetrahydronaphtalene (C1 and C4). C1 in green and C4 in orange. There are only represented mutants with values higher than PaDa-I (Mut0).



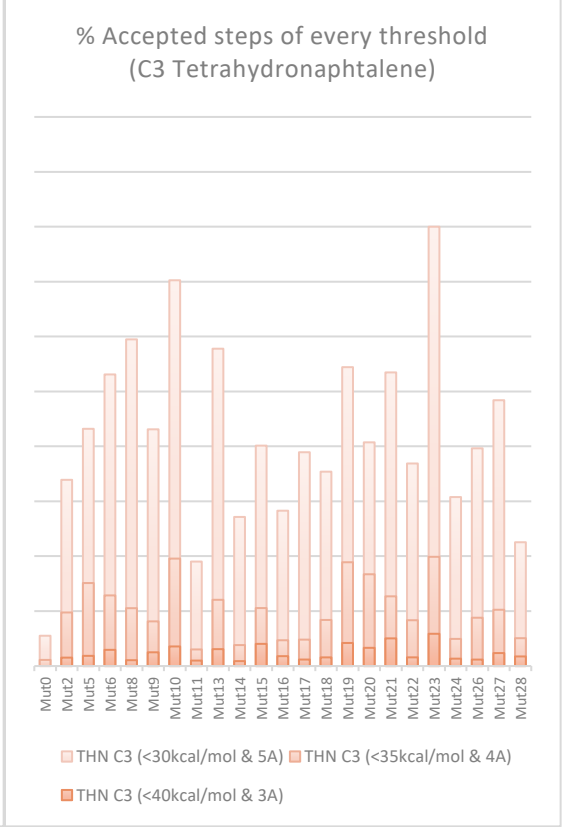


Figure 26 and Figure 27. Thresholds of the accepted steps with 1,2,3,4-tetrahydronaphtalene (C2 and C3). C1 in green and C4 in orange. There are only represented mutants with values higher than PaDa-I (Mut0).

5.2. Residence time

Another analysis that can be made from the simulations is the residence time. This parameter is different from the previous thresholds because not only considers the accepted but also the number of rejected Monte Carlo steps in which the substrate has stayed in a catalytic position. We considered that a substrate stays in a catalytic position when the distance between the oxo of the compound I and the carbon of interest of the substrate is under 5Å. This parameter is represented as a percentage, and it considers the total of steps (100 in our case).

The plot below is made from the bar diagrams found on the Supplemental material with the help of Sergi Rodà's scripts. On the plot there are only represented some values, the one's used in the analysis of the metric. The criteria used is exposed in the discussion section. There is one column for every mutant except in the case of substrates ethylbenzene (three carbons assessed) and 1,2,3,4-tetrahydronaphtalene (four carbons assessed) as seen before. The second case is because it is due to the symmetric carbons (C1 with C4 and C2 with C3).

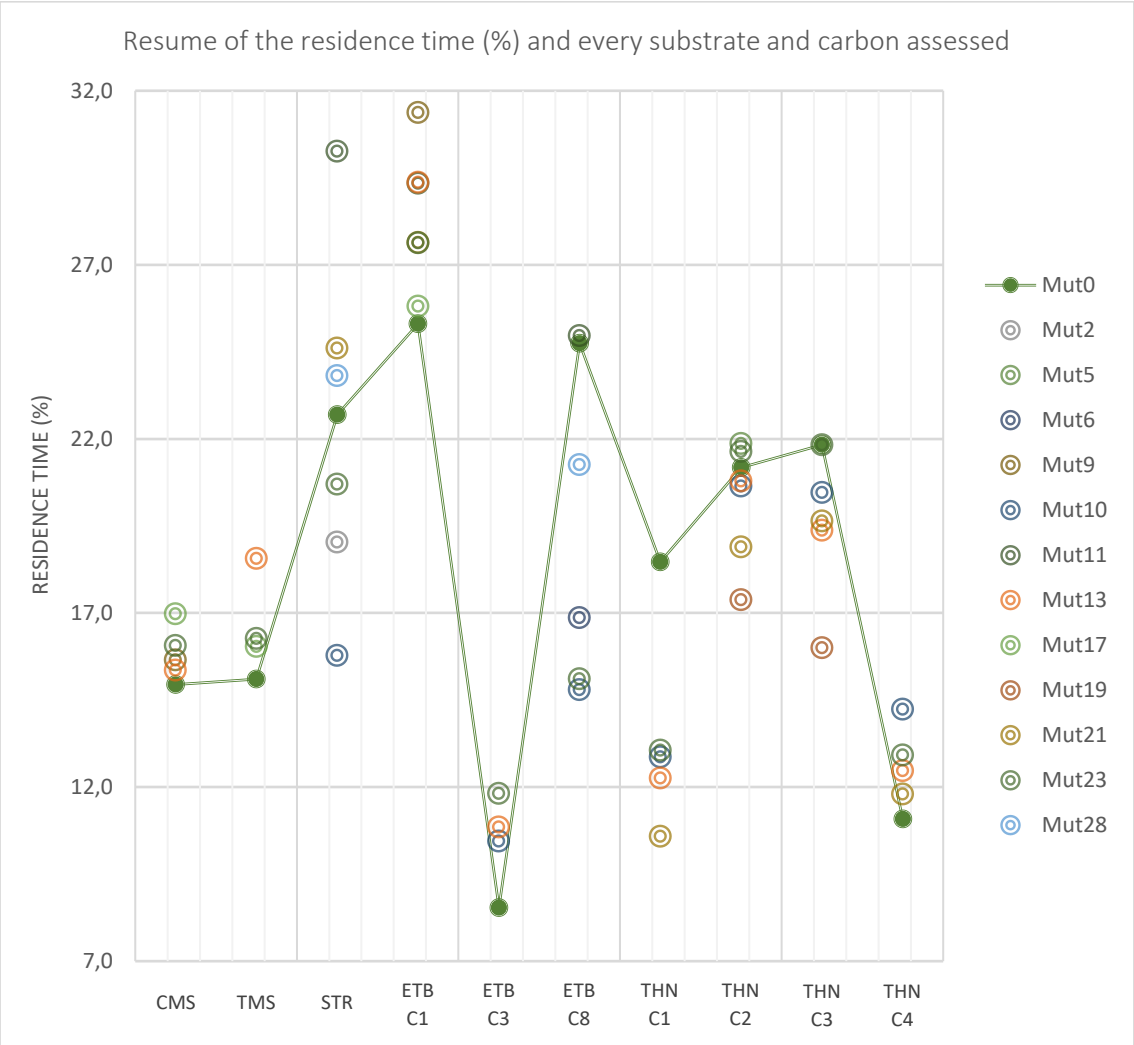


Figure 28. Residence time (% MC steps) under 5.0Å of mutants selected based on their percentage of accepted steps and best residence times (exposed in Table 1). The distances computed are the ones previously reported in Figure 16. Mutant 0 (PaDa-I) has a blue line to identify the threshold considered.

5.3. Best results for the combination of residence times and percentage of accepted steps

In *Table 1* there are marked in green the mutants selected as favorable candidates to consider for every substrate and position. The criteria used for this selection is exposed in the discussion section. Also, as said, are the only ones considered for the representation of the residence times in the plot.

Table 1. Resume of the results discussed. In green, mutants selected as notable to consider.

	CMS	TMS	STR	ETB (C1)	ETB (C3)	ETB (C8)	THN (C1/4)	THN (C2/3)
Mut 1								
Mut 2								
Mut 3								
Mut 4								
Mut 5								
Mut 6								
Mut 7								
Mut 8								
Mut 9								
Mut 10								
Mut 11								
Mut 12								
Mut 13								
Mut 14								
Mut 15								
Mut 16								
Mut 17								
Mut 18								
Mut 19								
Mut 20								
Mut 21								
Mut 22								
Mut 23								
Mut 24								
Mut 25								
Mut 26								
Mut 27								
Mut 28								
Mut 29								
Mut 30								

5.4. Enantioselectivity

Another aspect analyzed is the enantioselectivity of certain carbons of the substrates. These are the ones seen in *Figure 15.* We calculated the dihedral angle formed between the atoms highlighted in *Figure 30* in every case with the help of Sergi Rodà's scripts. These are the oxygen from the compound I, the carbon of interest, the adjacent carbon near the aromatic group and the other adjacent carbon. This dihedral was chosen according to a previous study in which asymmetric sulfoxidation of similar substrates was assessed[44]. Also, we followed the criteria of the same study to classify the angles. Every dihedral value was associated to a stablished category (*Figure 30,* right) depending on its value:

- From -40° to -140° as pro-S (a pose of the ligand that has more probability to become an S enantiomer).
- From 40° to 140° as pro-R.
- From -140° to 140° as undetermined because the poses could become with the same probabilities both pro-R and pro-S.
- From -40° to 40° as undetermined because the poses could become with the same probabilities both pro-R and pro-S or none.

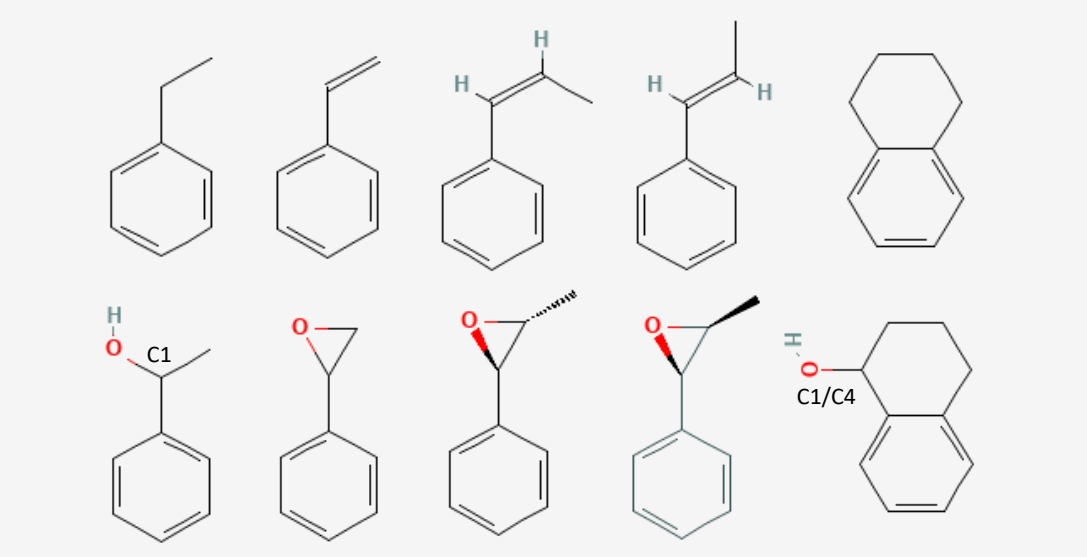


Figure 29. Positions in which the enantioselectivity is assessed. The substrates are above, and the products associated below. In ethylbenzene and 1,2,3,4-tetrahydronaphtalene the carbons are identified because in other analysis there are different carbons used.

Also, not all the poses were considered. We filtered them selecting only the ones below a catalytic distance of 5Å as reported previously[44]. Apart from this criterium, another one was added. The dihedrals were ranked by their energy binding from lowest to highest. From this classification, the top 50 dihedrals were selected, and the percentage of every enantiomer was calculated as can be seen in *Table 2.*

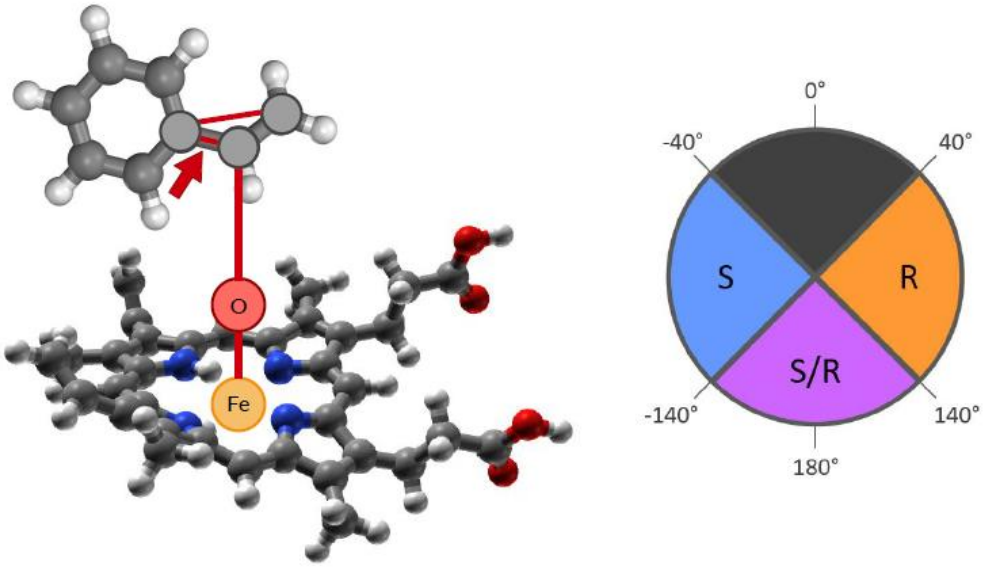
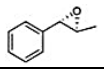
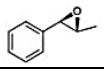
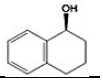


Figure 30. Dihedral angle used (red arrow) and angle ranges considered.

Table 2. Enantioselectivity of every mutant and substrate. Enantioselectivity is calculated from the dihedral angle of the accepted poses. It takes into account the 50 first poses with a distance < 5Å (catalytically relevant) ordered from higher to lower energies. The numbers below are the % of every enantiomer predicted. In green the results selected as promising (from Table 1) following the explanation of the discussion section.

	CMS		TMS		STR		ETB		THN	
	%1R2S	%1S2R	%1S2S	%1R2R	%R	%S	%R	%S	%R	%S
										
PaDa-I	63,64	36,36	64,52	35,48	52,17	47,83	81,48	18,52	70,37	29,63
Mut1	30,77	69,23	28,00	72,00	80,00	20,00	50,00	50,00	80,00	20,00
Mut2	50,00	50,00	53,33	46,67	83,33	16,67	46,15	53,85	44,44	55,56
Mut3	70,97	29,03	28,57	71,43	62,07	37,93	53,57	46,43	22,22	77,78
Mut4	0,00	100,00	100,00	0,00	0,00	100,00	36,36	63,64	63,64	36,36
Mut5	29,17	70,83	81,25	18,75	78,57	21,43	28,57	71,43	55,56	44,44
Mut6	64,00	36,00	55,88	44,12	28,00	72,00	37,14	62,86	18,18	81,82
Mut7	70,00	30,00	44,44	55,56	100,00	0,00	61,90	38,10	57,89	42,11
Mut8	28,95	71,05	42,11	57,89	50,00	50,00	78,13	21,88	55,26	44,74
Mut9	32,35	67,65	35,29	64,71	75,00	25,00	68,75	31,25	12,50	87,50
Mut10	65,00	35,00	56,25	43,75	100,00	0,00	51,72	48,28	58,82	41,18
Mut11	56,41	43,59	40,54	59,46	42,31	57,69	54,29	45,71	5,00	95,00
Mut12	58,33	41,67	69,57	30,43	42,86	57,14	46,67	53,33	57,14	42,86
Mut13	64,71	35,29	35,90	64,10	60,00	40,00	96,88	3,13	19,35	80,65
Mut14	65,38	34,62	15,63	84,38	36,36	63,64	82,14	17,86	0,00	100,00
Mut15	57,89	42,11	32,35	67,65	54,55	45,45	76,92	23,08	17,39	82,61
Mut16	34,38	65,63	68,89	31,11	50,00	50,00	88,24	11,76	6,67	93,33
Mut17	65,63	34,38	41,03	58,97	48,00	52,00	73,08	26,92	19,05	80,95
Mut18	63,33	36,67	27,78	72,22	45,45	54,55	97,67	2,33	6,90	93,10
Mut19	30,00	70,00	60,00	40,00	83,33	16,67	76,92	23,08	46,67	53,33
Mut20	65,00	35,00	65,00	35,00	55,56	44,44	65,63	34,38	17,39	82,61
Mut21	58,06	41,94	50,00	50,00	73,68	26,32	80,00	20,00	11,11	88,89
Mut22	64,71	35,29	46,67	53,33	41,67	58,33	93,33	6,67	60,87	39,13
Mut23	33,33	66,67	52,17	47,83	85,71	14,29	54,55	45,45	22,58	77,42
Mut24	35,29	64,71	27,27	72,73	45,45	54,55	86,67	13,33	4,35	95,65
Mut25	15,00	85,00	50,00	50,00	15,79	84,21	25,00	75,00	25,00	75,00
Mut26	44,00	56,00	40,00	60,00	43,75	56,25	85,71	14,29	39,13	60,87
Mut27	54,00	46,00	48,65	51,35	58,82	41,18	75,00	25,00	21,43	78,57
Mut28	64,10	35,90	48,65	51,35	57,14	42,86	57,14	42,86	20,00	80,00
Mut29	50,00	50,00	40,00	60,00	53,85	46,15	52,94	47,06	31,03	68,97
Mut30	20,00	80,00	81,48	18,52	36,00	64,00	32,14	67,86	88,89	11,11

6. Discussion

6.1. Analysis of the binding energies and distances

The results will be first analyzed by substrate, to identify which mutants have the best values for every ligand. For every of them, the two metrics assessed will be compared between substrates. We must mention that the identity of the mutants is blind because that information is subject to be on a publication and it is confidential.

For the *cis-β*-methylstyrene substrate, in terms of the percentage of accepted steps in the biggest threshold (*Figure 18*), we observe that mutants 9, 13, 17 and 23 have remarkably good values (in special Mut23 because it has values around 14%). This is partially supported by the residence time analysis (*Figure 28*) because the mutants previously highlighted are among the best percentages. We must mention that, apart from these four, mutants 11 and 24 have also higher residence times than PaDa-I and, in the case of the 11, than all the mutants. If we look for the plot of this mutant, we can see that the density of dots is higher than the one of PaDa-I, but it is also a mediocre approach to the oxygen of the heme compared to other variants because its percentage is around 8%. Because of this, we cannot consider it a good candidate to take into account. To sum up, the best candidates are 23, 17, 13 and 9 in that order from the most to the least significant. We consider 17 before 13 because although it has a lower density of dots in the bigger threshold (under -30kcal/mol and 5Å), has almost a 4% of accepted steps in the second one (under -35kcal/mol and 4Å) and a higher residence time (around a 17%).

For *trans*-β-methylstyrene, the results are more dispersed between the two analyses. In terms of residence time (*Figure 28*), mutants 9, 11, 13, 17, 21, 23, 24, 28 and 30 have good values around 16-18%. From all of these, variants 11, 13 and 30 are the best three. Regarding to the thresholds (*Figure 19*), mutants 23, 13, 8, 17 and 2 have a right density of dots (from highest to lowest) although 23 is the only one that reaches a 14% in the biggest threshold. Intercrossing the two analyses, only variants 13, 17 and 23 show proper values for both. Between the 13 and 17 structures, we see that mutant 13 has better residence time, and density of dots. We cannot stablish a ranking between 13 and 23 and both should be considered.

In the case of the residence time of styrene, only mutants 11, 21 and 28 have better results than PaDa-I in residence time (*Figure 28*). However, we have to mention that although they are the highest values, all residence times of styrene are very high (around 20%). As we are looking for variants that could be better than PaDa-I, we only consider the ones with higher values than it. Because of this, mutants 11, 21 and 28 have good residence times, reaching a 30%, 25% and 24%, respectively. Variants 11, 19, 21 and 23 have good percentages in their thresholds (*Figure 20*), specially the later with values of almost 25% and 27%, respectively in the bigger threshold. Also, we must consider mutants 2, 10 and 28 in the analysis of the thresholds. Although they have a poor density in the bigger threshold, the second (under -35kcal/mol and 4Å) and third ones (under -40kcal/mol and 5Å) exhibit higher values than the ones shown by other mutants. For

example, mutants 10 and 28 show a 1% of accepted steps in the smaller threshold, similar values to the ones of mutant 23. In summary, variants 21, 23 and 28 are the main to consider but also 2, 10, and 11 could result in good outcomes.

If we analyze the three interesting points of ethylbenzene (carbon 1, orto, and para positions), we can see clearly more favorable results for some carbons. The comparative between the different positions will be commented in the next section (6.2. Analysis of the regioselectivity on ethylbenzene and 1,2,3,4-tetrahydronaphtalene).

Analyzing them separately and going first with the **carbon 1**, we can see that in terms of time of residence (*Figure 28*), mutants 9, 11, 13, 17, 21, 23 and 24 have higher residence times than PaDa-I. Its values vary between 23-30%. If we take a look at the thresholds (*Figure 21*), mutants 8, 9, 10, 11, 13 and 23 have proper density of dots (from about 27-30% and almost 40% for the latest). We could also add both 17 and 21 mutants for its good results in the second range (about 12%). Between 8, 9, 10, 11, 17 and 21, we see that the first has a poorer density at the second threshold and that the other five could be considered quite similar. Mutant 13 and, remarkably, 23 have the best results in this analysis. If we cross these results with the residence time, mutants 13 and 23 have great values and 9, 11, 17 and 21 also are higher than the reference.

If we take a look at the **orto position** (the one situated at the carbon 3), we can see that it has pretty low residence times (*Figure 28*) but many of the mutants have higher density of dots than PaDa-I in the other analysis. Considering this and as an attempt to correlate the two values, we see that mutants 10, 13, 19, 20 and 23 have an interesting density of dots (*Figure 22*), (going around 26% for 10 and 13, 21% for 19 and 20 and 35% for the 23). Mutants 10, 13 and 23 are the ones with higher values in the first range. Although 19 and 20 don't seem to have the best profiles, they show a high percentage in the second and third ranges (about a 5% in the second range). If we take a look to the residence times, among the mutants that show (very little) difference regarding to PaDa-I are 5, 10, 11, 13, 20, 23, 24 and 28. The ones that can be correlated to the anterior analysis are 10, 13 and 23. To sum up and with both analyses in hand, mutants 10, 13 and 23 seem that could result in good outcomes.

Following with the last carbon to evaluate from ethylbenzene, the **para position** (the one at the carbon 8), we can see a clear difference between this one and the other positions assessed. This time, residence times (*Figure 28*) are less ambiguous because there is only one mutant with a higher residence time than PaDa-I, mutant 11 (25%). Looking at the bars (*Figure 23*) (although with lower percentages than the other two), mutants 6, 10, 11, 15, 23, and 28, especially the latter two show the best percentages (around a 15%). For this, we will consider mutants 6, 10, 11, 23 and 28.

The substrate 1,2,3,4-tetrahydronaphtalene show very similar results in the four positions in the analysis of the percentage of accepted steps. At the residence time we

see more biased results. As with the case of ethylbenzene, the results between positions will be analyzed in the next section (6.2. Analysis of the regioselectivity on ethylbenzene and 1,2,3,4-tetrahydronaphtalene).

Observing the results for the **C1/C4 pair**, the results for the residence times (*Figure 28*) are quite despaired but with a general view only mutants 2, 5 and 10 show greater values than PaDa-I depending on the carbon (1 or 4) observed. We see that variant 5 has a percentage of around 20% (the highest) while mutants 2 and 10 have values of around 13-14% comparing the numbers to carbon 4. For the analysis of the density of dots (*Figures 24 and 25*), almost all the mutants exhibit greater values than the mutant 0. Variants 5, 10, 13, 21 and, specially, 23 (with a value of 18%) are the ones highlighted. Also, we have to mention that the percentages of the second threshold are very high (around a 6-7%).

Switching to the **C2/C3 pair**, the only values that surpass the ones from PaDa-I on the residence time (*Figure 28*) analysis are the ones of mutants 5, 11 and 23 (with values around 21%). Looking at the thresholds (*Figures 26 and 27*), almost all the mutants show higher values than mutant 0. We can stand out mutants 5, 10, 13, 19, 21 and 23 (the latest with a 16-18%). Correlating both analyses and summarizing the list, mutants 5, 10, 13, 19, 21 and, remarkably, 23 have good values.

In the mutants considered above as rellevant, the results with good residence times are almost always considered. Although some mutants don't have ideal values of residence time, we will consider some of them based on their thresholds. All the results commented are summarized in *Table 1. F*or *cis*-β-methylstyrene, interesting mutants are 9, 13, 17 and 23 with high percentages on the residence times and percentages of accepted steps of 12.4% (Mut9), 13.3% (Mut13), 13.1% (Mut17), and 13.7% (Mut23) for the biggest threshold. For *trans*-β-methylstyrene, also mutants 13, 17 and 23 show great values at the residence times and significant percentage of accepted steps of 10.8% (Mut13), 9.9% (Mut17), and 12.8% (Mut23), for the biggest threshold. The similarity between the results of these two substrates can be explained because they are very alike and they could migrate to the active site in an almost identical way. However, CMS seem to have slightly better values. In the case of styrene, mutants 21 and 28 are the principal ones but 2, 10, 11, and 23 should be also considered. Considering the previous mutants, their percentage of accepted steps in the bigger threshold are 19.0% (Mut2), 17.5% (Mut10), 20.0% (Mut11), 24.0% (Mut21), 26.8% (Mut23), and 16.3% (Mut28). We have to mention that styrene has the highest values of residence time (reaching ~24% and even higher values in some cases). For ethylbenzene and its carbon 1, mutants show values of 24.5% (Mut9), 23.3% (Mut11), 30.8% (Mut13), 21.6% (Mut17), 21.0% (Mut21), and 39.3% (Mut23). Also, we see that C1 has very good values of residence time (~23%). Here, we can see that some mutants retrieved for styrene and ethylbenzene are similar too as well as the residence time (the two best). Considering that the molecule is almost identical and the carbon attacked is almost the same (in the case of ethylbenzene, the carbon is attacked and in styrene it is the double bond because it outputs an

epoxidation) it is coherent. Also, the values for both styrene and ethylbenzene are higher than the ones of CMS and TMS. This is probably due to their size because CMS and TMS are bigger substrates and it could be more difficult for them to move inside the cavity. Taking a look into the best mutants and their percentage of accepted steps in the bigger threshold, we can see that for orto are 26.6% (Mut10), 27.0% (Mut13) and 34.8% (Mut23), and for para are 10.4% (Mut6), 16.8% (Mut10), 11.0% (Mut11), 22.0% (Mut23), and 11.0% (Mut28). Comparing these results with the ones obtained by the same substrate but in carbon 1, we can see that this latter has better values in general, especially in the second threshold. This could be because orto and para positions are in the aromatic group, bigger than the apex in where is situated carbon 1, and thus, easier to be orientated towards the heme. Finally, for 1,2,3,4-tetrahydronaphtalene at the C1/C4 pair, best mutants and their percentages of accepted steps in the bigger threshold are 8.6/9.4% (Mut5), 16.8/13.6% (Mut10), 13.8/10.7% (Mut13), 12.0/10.1% (Mut21), and 18.6/15.7% (Mut23). Also, for C2/C3 are 9.0/8.6% (Mut5), 16.6/14.1% (Mut10), 13.5/11.6% (Mut13), 13.2/10.9% (Mut19), 12.0/10.7% (Mut21), and 18.6/16.0% (Mut23). Between the 2 paired positions of tetrahydronaphtalene, we can see that both have very similar good mutants. Compared with other substrates, tetrahydronaphtalene has values comparable to the ones of *cis*- and *trans*-*β*-methylstyrene or a little bit higher. Maybe this can be confusing because 1,2,3,4-tetrahydronaphtalene is a bigger substrate than the previous two and should be more difficult to be moved. Results don't show this tendency because both results are quite similar (even tetrahydronaphtalene has better percentages in the second thresholds).

A wider analysis about the differences in the same substrate with different carbons will be done in next section as said before. As a general view, it is clear that three mutants have represented good results in a general manner. This are mutants 10, 13 and 23. They have been in five, six and eight of the positions assessed, respectively. As stated in section *"5.3. Best results for the combination of residence times and percentage of accepted steps"*, *Table 1* resumes the best results commented.

6.2. Analysis of the regioselectivity on ethylbenzene and 1,2,3,4-tetrahydronaphtalene

This analysis aims to analyze the preference of the mutants for certain positions over all other.

6.2.1. Ethylbenzene

First of all, the distances used for both analyses in the orto and para positions are computed as epoxidations. This is due to the hydroxylation is proposed to occur via an epoxidation[10]. If we take a look at both analyses, the orto position (C3) has better values of the percentage of accepted steps, and contrary to this, the para position (C8) has higher general results in the residence time. Following with the residence time plot (*Figure 28*), although the para position has higher general values (of around 15-20%) of residence time in the mutants, the value for PaDa-I is very elevated too (25%). This makes that only one mutant (Mut11) lightly surpasses the reference. Switching to the orto position, we see that in general values are found approximately at the 10%. Even

though the values are low, many mutants have higher percentages than PaDa-I (8.5%) and can be promising for its own activities. As we are comparing the preference for the two positions, better values of residence time could represent a higher affinity (for the para position in this case). If we talk about the other analysis, orto position is the one outlined now. It has better values mainly in the bigger (around 15-25%) and in the second threshold. In this case, para position has values of around 10-15% in the bigger threshold.

The two results are incongruous and *a priori* we cannot consider an analysis better than the other. It is true that a higher residence time indicates that the substrate has stayed in a catalytic position (under 5Å) during a greater percentage regarding to the total of steps (accepted and not accepted). It computes the steps that the substrate has stayed in a nearby pose to the heme until a new step is accepted (and depending on the distance to the heme of this new step, this will be computed also or not). However, the percentage of accepted steps in every threshold informs us of the quality of the accepted steps of the simulation. 3Å to the heme is a distance more relevant than a one of 5Å and because of this, the orto position seems to have a "better quality" of the catalytically relevant poses as seen. Maybe an analysis of another residence time taking a smaller range (4Å for example) could bring us more information about this dichotomy. In conclusion, orto position has less percentage of residence time but its accepted steps are of higher quality (closer to the heme) than para position.

Trying to stablish mutants with preference for only one position of the two. In the case of a mutant with more preference for the orto position (C3), only one has better residence time compared to the para position. This is mutant 30 with values of about 9.8% (orto) and 8.4 (para). If we look to the percentage of dots, mutant 30 isn't in the *Figure 22* due to their poorer results compared to PaDa-I. Because of this we cannot state that mutant 30 is a good variant to consider for the hydroxylation of the orto position. However, all the other mutants have higher values in the orto position. If we switch the analysis, looking for a mutant with higher values of residence time for para position is "easier" although there's only one mutant with higher values than PaDa-I and that its orto position. This is mutant 11. Nonetheless, in the analysis of the thresholds of this mutant, we see that although it has poorer values in the bigger threshold, the second one is like the values obtained by the para position and in the third threshold is even higher.

To sum up, in general all the mutants show to have lower residence times for the orto position but with closer poses and higher energies. So, there are less probabilities that a substrate stays in a catalytic position but when this happens, the interactions will be more efficient. The opposite happens with the para position, there are more possibilities that a substrate stays in a catalytic pose but when this occurs, the interactions seem to be weaker.

6.2.2. 1,2,3,4-tetrahydronaphtalene

This time the distances computed aren't considered as epoxidations because the reaction is performed as a direct hydroxylation. As stated previously the 2 positions assessed (that can retrieve the compounds as 1-ol and 2-ol) are double because of the symmetry of the molecule. Because of this, carbons 1 and 4 are considered for the formation of the 1-ol and carbons 2 and 3 are considered for the formation of the 2-ol. With the double of results, depending on the initial perturbations and the path followed by every report, the outcomes can be very diverse. If we look at the residence times, we see a clear tendency in general for carbons 1 and 4 to have lower percentages than the carbon 2/3 pair. However, if we compare the results to PaDa-I, very few can surpass its values and, therefore, be better than the reference. One of these mutants is mutant 5 for both pairs (C1/C4 and C2/3) with values of 21.9/20.6% (C1/C4) and 23.5/24.7% (C2/C3). Another one is mutant 11 with higher values only for the C2/C3 pair, having a 22.5/21.4% (C2/C3). There is only one more mutant with higher values than the reference but only in the C2/C3 pair, mutant 23. This mutant has values of 21.6/21.8% (C2/C3).

Looking at the percentage of accepted steps, the values for all 4 carbons assessed are very similar by pairs. In general, C1 and C2 plots seem to have higher results than the ones of C3 and C4. We cannot point at specific mutants with high affinity for one carbon and low for the other because all the results are alike.

In summary, it is true that the pair C2/C3 has better residence times. As the other results are almost the same, a higher residence time could guarantee a major probability of higher binding energies and because of this this pair shows better results. This could be because of the position of the carbons; it could be easier for this pair to be orientated towards the heme.

6.3. Analysis of the enantioselectivity

This analysis aims to analyze the preference of the mutants for a certain enantiomer over the other one. Although there are mutants very polarized for one enantiomer, their residence times and thresholds must be considered too. An interesting ratio means almost nothing if that enzyme does not retrieve high product formation yields (and thus, its values of residence time and percentage of accepted steps must be considered).

If we take a look to *Table 1,* in green are marked those mutants mentioned on the summary of the analysis of the binding energies and distances. On the first line there are the values obtained for the enantioselectivity of PaDa-I. In the case of CMS all the values highlighted can be considered as they are ratios of about 60/30. Comparing the results with the ratios, we can see that mutants 13 and 17 maintain the ratio of PaDa-I (about 60R/30S) and mutants 9 and 23 invert the enantioselectivity (30R/60S). Switching to TMS, mutant 13 is the only one that shows that is predilect for one of the enantiomers as it has values of 35.9S/64.1R. Compared to PaDa-I, mutant 13 is inverting the enantioselectivity of the reference. Mutants 17 and 23 have ratios close to the 50/50. In STR, mutants 2 (83.3R/16.7S), 10 (100.0R/0S), 21 (73.7R/26.3S) and 23 (85.7R/14.3S)

have clear selectivity for one enantiomer. Also, mutants 11 and 28 show percentages near to the 50/50 as the original mutant. For ETB, mutants 9 (68.8R/31.2S), 13 (96.9R/3.1S), 17 (73.1R/26.9S) and 21 (80.0R/20.0S) have polarized percentages while mutants 11 and 23 are closer to the 50/50. In this case, PaDa-I shows preference for the R enantiomer as well as the mutants with preference for one of the enantiomers. Finally, in THN, mutants 13 (19.4R/80.6S), 21 (11.1R/88.9S) and 23 (22.6R/77.4S) are focused on one enantiomer. Mutants 5, 10 and 19 are near the 50/50. In this example, PaDa-I shows preference for the R enantiomer, contrary to the values of the rest of the mutants assessed above. Maybe mutant 10 (58.8R/41.2S) could be considered although these values close to the 50/50.

In general, we can observe that almost all the substrates have very diverse values for both enantiomers. If we take a look to THN, nearly all the mutants have preference for one enantiomer (S). This could be due to the size of the molecule. If the initial pose used to run the simulations is a pro-S pose it is more probable that that substrate will stay in a similar orientation as bigger is the substrate. As this could affect to the results, larger alternative simulations should be carried out maybe with stronger perturbations to force the possible flip of the molecule.

The simulations used for all the analysis are the same. For next analysis it should be studied to perform longer simulations (of about 500 steps) of only one epoch. The single epoch is due to the limitations of AdaptivePELE where at each epoch change the Boltzmann sampling is broken, and less favored areas are chosen for the next exploration. The simulation moves to investigate all the possible areas making it more difficult to converge population analysis.

7. Conclusions

The observed enzyme and substrate interactions were rationalized by computational simulations. Both enzyme and substrate structural determinants rule the observed activities, regio- and enantioselectivities. Such determinants, related to the different mutations of every variant and the structure and size of the different substrates, need to be taken into account for developing improved UPOs' variants.

In the evaluation of PaDa-I's mutants in terms of binding energy, the two analyses helped us to examine the results obtained in the energy plots from two perspectives. We have obtained very different results and very promising mutants with high values of both parameters for the different substrates. This proves that PELE can identify differences between the mutants and the ligands. The effect of the ligands it easier to analyze considering their size and shape. As smaller the substrate, more mobility it has inside the cavity and greater values retrieve in the two analyses. However, this is not always true because mutations play a giant roll. Also, differences were observed when assessing diverse carbons on the same substrate. The placing of the substrates by PELE can tell us the preference of every mutant for a certain carbon.

In addition, simulations showed that the enantioselectivity of the variants can be deducted from the orientation taken in the active site: the dihedral angle formed. This means that is a variable parameter that depends on the evolution of the simulation towards the preference of every substrate and mutant combination. The starting point is the same for all the substrates because they are the same poses. Because of this, the substrates begin with a 50/50 of probabilities to turn to one enantiomer or the other. The fact that we obtained results very polarized for the two enantiomers with the same substrate means that they have had to perform a flip and had to stay in that determinate angle no matter what depending on the environment of the mutant.

We realized that in some cases (especially in the ones that involved THN) results could be biased towards one enantiomer. This could be due to the incapacity of the substrate to move freely inside the cavity, possibly explained considering that THN is the bigger substrate of the five. If the initial pose used to run the simulation is a pro-S pose or if the initial perturbation places it like this, it is more probable that a bigger substrate stays in a similar orientation. Because of this, larger, alternative simulations without the AdaptivePELE module should be carried out with stronger perturbations to force the possibility molecule flipping.

It is notable to mention that, in contrast to the high number of studies on P450 and other enzymes engineering[6,12], only a few publications on UPO and directed evolution[1,24,35,36,44,45], rational design or computational analysis have been published to date. Additional studies combining both experimental and computational approaches will stablish a strong base to the ideal of redesigning UPO biocatalysts in compounds of interest. In fact, this study is going to be compared with experimental data to complete the rationalization of the results. The increasing interest on UPOs id demonstrated by the high number of review articles published[6,12,19,39,46].

8. Bibliography

1. Linde, D. *et al.* Two new unspecific peroxygenases from heterologous expression of fungal genes in Escherichia coli. *Appl. Environ. Microbiol.* **86**, 1–16 (2020).

2. Ramirez-Escudero, M. *et al.* Structural Insights into the Substrate Promiscuity of a Laboratory-Evolved Peroxygenase. *ACS Chem. Biol.* **13**, 3259–3268 (2018).

3. R, U., J, N., K, S., J, S. & M, H. Novel haloperoxidase from the agaric basidiomycete Agrocybe aegerita oxidizes aryl alcohols and aldehydes. *Appl. Environ. Microbiol.* **70**, 4575–4581 (2004).

4. Ullrich, R. & Hofrichter, M. The haloperoxidase of the agaric fungus Agrocybe aegerita hydroxylates toluene and naphthalene. *FEBS Lett.* **579**, 6247–6250 (2005).

5. Bormann, S. *et al.* Modeling and simulation-based design of electroenzymatic batch processes catalyzed by unspecific peroxygenase from A. aegerita. *Biotechnol. Bioeng.* **118**, 7–16 (2021).

6. R, U. & M, H. Enzymatic hydroxylation of aromatic compounds. *Cell. Mol. Life Sci.* **64**, 271–293 (2007).

7. Hofrichter, M. & Ullrich, R. New Trends in Fungal Biooxidation. *Ind. Appl.* 425–449 (2011) doi:10.1007/978-3-642-11458-8_21.

8. MG, K., R, U., K, S. & M, H. Spectrophotometric assay for detection of aromatic hydroxylation catalyzed by fungal haloperoxidase-peroxygenase. *Appl. Microbiol. Biotechnol.* **75**, 1473–1478 (2007).

9. R, U., C, D., M, K. & M, H. Pyridine as novel substrate for regioselective oxygenation with aromatic peroxygenase from Agrocybe aegerita. *FEBS Lett.* **582**, 4100–4106 (2008).

10. Peter, S., Kinne, M., Ullrich, R., Kayser, G. & Hofrichter, M. Epoxidation of linear, branched and cyclic alkenes catalyzed by unspecific peroxygenase. *Enzyme Microb. Technol.* **52**, 370–376 (2013).

11. Hofrichter, M., Kellner, H., Pecyna, M. J. & Ullrich, R. Fungal unspecific peroxygenases: Heme-thiolate proteins that combine peroxidase and cytochrome P450 properties. *Adv. Exp. Med. Biol.* **851**, 341–368 (2015).

12. Hofrichter, M., Ullrich, R., Pecyna, M. J., Liers, C. & Lundell, T. New and classic families of secreted fungal heme peroxidases. *Appl. Microbiol. Biotechnol.* **87**, 871–897 (2010).

13. Hofrichter, M. & Ullrich, R. Heme-thiolate haloperoxidases: versatile biocatalysts with biotechnological and environmental significance. *Appl. Microbiol. Biotechnol.* **71**, 276–288 (2006).

14. Faiza, M., Huang, S., Lan, D. & Wang, Y. New insights on unspecific peroxygenases: Superfamily reclassification and evolution. *BMC Evol. Biol.* **19**, 1–19 (2019).

15. Kellner, H. *et al.* Widespread Occurrence of Expressed Fungal Secretory Peroxidases in Forest Soils. *PLoS One* **9**, e95557 (2014).

16. Pecyna, M. J. *et al.* Molecular characterization of aromatic peroxygenase from Agrocybe aegerita. *Appl. Microbiol. Biotechnol.* **84**, 885–897 (2009).

17. Gupta, D. K. *et al.* The genome sequence of the commercially cultivated mushroom Agrocybe aegerita reveals a conserved repertoire of fruiting-related genes and a versatile suite of biopolymer-degrading enzymes. *BMC Genomics* **19**, 1–13 (2018).

18. Wang, X., Ullrich, R., Hofrichter, M. & Groves, J. T. Heme-thiolate ferryl of aromatic peroxygenase is basic and reactive. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3686–3691 (2015).

19. Hofrichter, M. & Ullrich, R. Oxidations catalyzed by fungal peroxygenases. *Curr. Opin. Chem. Biol.* **19**, 116–125 (2014).

20. Pabulo H. Rampelotto. *Grand Challenges in Fungal Biotechnology. Chapter 12 Spotlight on Class I Hydrophobins: Their Intriguing Biochemical Properties and Industrial Prospects*. (2020).

21. Piontek, K. *et al.* Structural basis of substrate conversion in a new aromatic peroxygenase: Cytochrome P450 functionality with benefits. *J. Biol. Chem.* **288**, 34767–34776 (2013).

22. Martin-Diaz, J., Molina-Espeja, P., Hofrichter, M., Hollmann, F. & Alcalde, M. Directed evolution of unspecific peroxygenase in organic solvents. *Biotechnol. Bioeng.* **118**, 3002–3014 (2021).

23. Ramirez-Ramirez, J., Martin-Diaz, J., Pastor, N., Alcalde, M. & Ayala, M. Exploring the role of phenylalanine residues in modulating the flexibility and topography of the active site in the peroxygenase variant pada-i. *Int. J. Mol. Sci.* **21**, 1–15 (2020).

24. Molina-Espeja, P. Diseño de la peroxigenasa inespecífica de Agrocybe aegerita mediante evolución dirigida: expresión funcional en levaduras y síntesis de 1-naftol. (2016).

25. Kenneth W. Borrelli, Andreas Vitalis, Raul Alcantara, and & Guallar*, V. PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J. Chem. Theory Comput.* **1**, 1304–1311 (2005).

26. Madadkar-Sobhani, A. & Guallar, V. PELE web server: atomistic study of biomolecular systems at your fingertips. *Nucleic Acids Res.* **41**, W322–W328 (2013).

27. Lecina, D., Gilabert, J. F. & Guallar, V. Adaptive simulations, towards interactive protein-ligand modeling. *Sci. Reports 2017 71* **7**, 1–11 (2017).

28. Lecina-Casas, D., Takahashi, R. & Guallar, V. On the way to real time protein-ligand sampling.

29. Hernández-Ortega, A. *et al.* Substrate diffusion and oxidation in GMC oxidoreductases: an experimental and computational study on fungal aryl-alcohol oxidase. *Biochem. J.* **436**, 341–350 (2011).

30. Hosseini, A. *et al.* Molecular Interactions of Prodiginines with the BH3 Domain of Anti-Apoptotic Bcl-2 Family Members. *PLoS One* **8**, e57562 (2013).

31. Espona-Fiedler, M. *et al.* Identification of dual mTORC1 and mTORC2 inhibitors in melanoma cells: Prodigiosin vs. obatoclax. *Biochem. Pharmacol.* **83**, 489–496 (2012).

32. Borrelli, K. W., Cossins, B. & Guallar, V. Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *J. Comput. Chem.* **31**, 1224–1235 (2010).

33. Lucas, M. F. & Guallar, V. Single vs. multiple ligand pathways in globins: A computational view. *Biochim. Biophys. Acta - Proteins Proteomics* **1834**, 1739–1743 (2013).

34. Lucas, M. F. & Guallar, V. An Atomistic View on Human Hemoglobin Carbon Monoxide Migration Processes. *Biophys. J.* **102**, 887–896 (2012).

35. Mate, D. M., Palomino, M. A., Molina-Espeja, P., Martin-Diaz, J. & Alcalde, M. Modification of the peroxygenative: Peroxidative activity ratio in the unspecific peroxygenase from Agrocybe aegerita by structure-guided evolution. *Protein Eng. Des. Sel.* **30**, 191–198 (2017).

36. Molina-Espeja, P. *et al.* Synthesis of 1-Naphthol by a Natural Peroxygenase Engineered by Directed Evolution. *ChemBioChem* **17**, 341–349 (2016).

37. Municoy, M. *et al.* Fatty-Acid Oxygenation by Fungal Peroxygenases: From Computational Simulations to Preparative Regio: From Stereoselective Epoxidation. *ACS Catal.* **10**, 13584–13595 (2020).

38. Babot, E. D. *et al.* Steroid hydroxylation by basidiomycete peroxygenases: A combined experimental and computational Study. *Appl. Environ. Microbiol.* **81**, 4130–4142 (2015).

39. Martínez, A. T. *et al.* Oxidoreductases on their way to industrial biotransformations. *Biotechnol. Adv.* **35**, 815–831 (2017).

40. Kiebist, J. *et al.* One-pot synthesis of human metabolites of SAR548304 by fungal peroxygenases. *Bioorg. Med. Chem.* **23**, 4324–4332 (2015).

41. Kinne, M. *et al.* Oxidative Cleavage of Diverse Ethers by an Extracellular Fungal Peroxygenase *. *J. Biol. Chem.* **284**, 29343–29349 (2009).

42. Olmedo, A. *et al.* From Alkanes to Carboxylic Acids: Terminal Oxygenation by a Fungal Peroxygenase. *Angew. Chemie Int. Ed.* **55**, 12248–12251 (2016).

43. Lucas, F. *et al.* Molecular determinants for selective C 25-hydroxylation of vitamins D 2 and D 3 by fungal peroxygenases.

44. Linde, D. *et al.* Asymmetric sulfoxidation by engineering the heme pocket of a dye-decolorizing peroxidase. *Catal. Sci. Technol.* **6**, 6277–6285 (2016).

45. Gomez De Santos, P. *et al.* Selective Synthesis of the Human Drug Metabolite 5'-Hydroxypropranolol by an Evolved Self-Sufficient Peroxygenase. *ACS Catal.* **8**, 4789–4799 (2018).

46. Sigmund, M.-C. & Poelarends, G. J. Current state and future perspectives of engineered and artificial peroxygenases for the oxyfunctionalization of organic molecules. *Nat. Catal. 2020 39* **3**, 690–702 (2020).

9. Abbreviations

***Aae*UPO**	Unspecific Peroxygenase from Agrocybe a*egerita*
ANM	Anisotropic Normal Mode
APO	Aromatic Peroxygenase
C0	Compound 0
CI	Compound I
CII	Compound II
***Cci*UPO**	Unspecific Peroxygenase from *Coprinopsis cinerea*
CMS	*Cis*-β-methylbenzene
CPO	Chloroperoxidases
***Cra*UPO**	Unspecific Peroxygenase from *Coprinellus radians*
DyP	Dye decolorizing peroxidase
ETB	Ethylbenzene
FAP	Fatty acid peroxygenase
MC	Monte Carlo
MD	Molecular Dynamics
MP	Myeloperoxidase
***Mro*UPO**	Unspecific Peroxygenase from *Marasmius rotula*
NADH-POD	NADH peroxidase - POD (phenol oxidizing)
P450	Cytochrome P450 monooxygenase
PCP	Proline-Cysteine-Proline
PELE	Protein Energy Landscape Exploration
PSP	Plant seed peroxygenase
QM/MM	Quantum mechanics/molecular mechanics
STR	Styrene
THN	1,2,3,4-tetrahydronaphtalene
TMS	*Trans*-β-methylbenzene
UPO	Unspecific Peroxygenase
VP	Versatile peroxidase

10. Acknowledgements

I would like to express my gratitude to Dr. Victor Guallar Tasies for giving me his guidance, valuable advice and experience during my internship and giving me the opportunity to join to the Electronic and Atomic Protein Modelling Research Group. Also, I would also like to extend my thanks to Sergi Rodà Llordes for their patience, advice and help along the way.

In addition, I would like to thank my family and friends for they support. You are always there for me.

11. Supplemental material

11.1. Thresholds of all the mutants

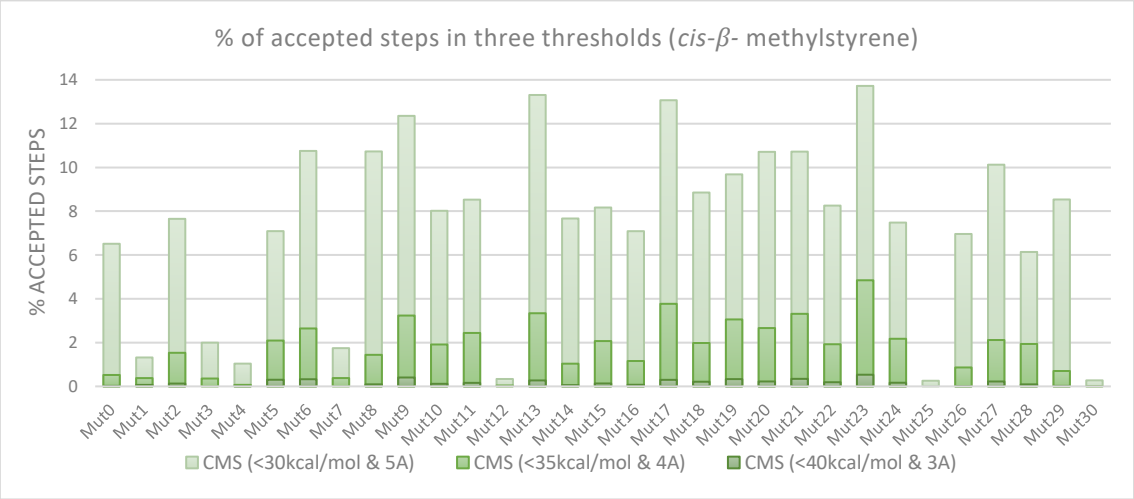


Figure 31. Thresholds of the accepted steps with cis-β-methylstyrene. All mutants are represented.

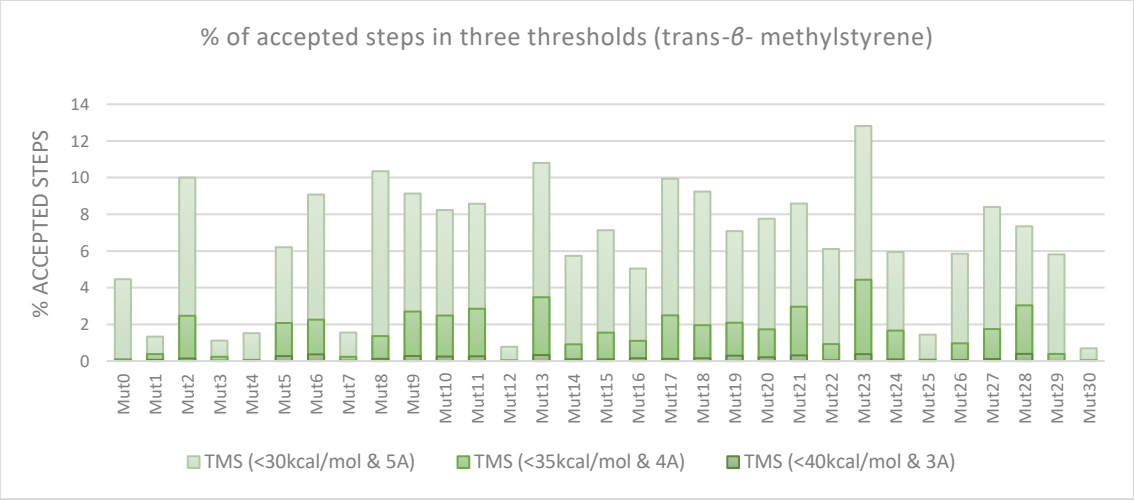


Figure 32. Thresholds of the accepted steps with trans-β-methylstyrene. All mutants are represented.

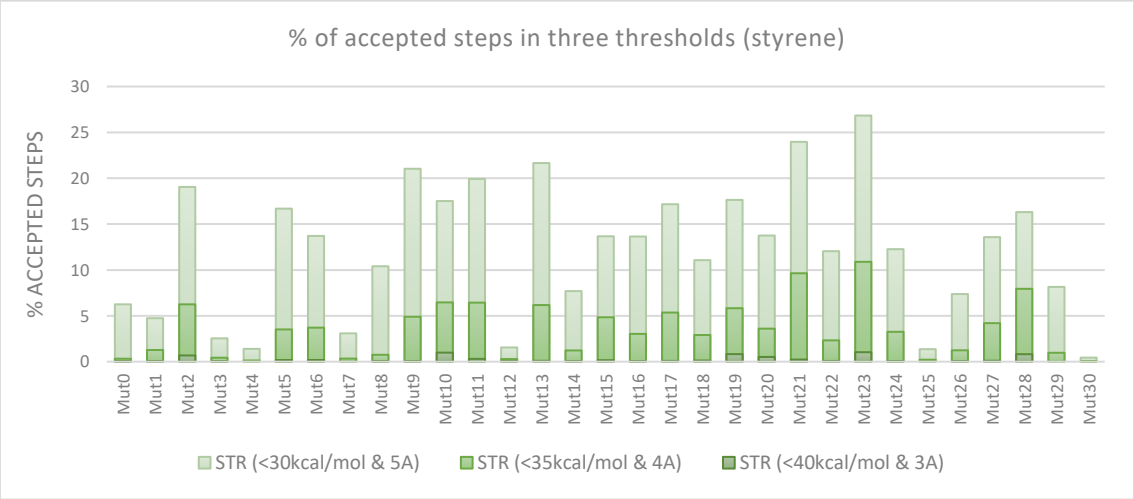


Figure 33. Thresholds of the accepted steps with styrene. All mutants are represented.

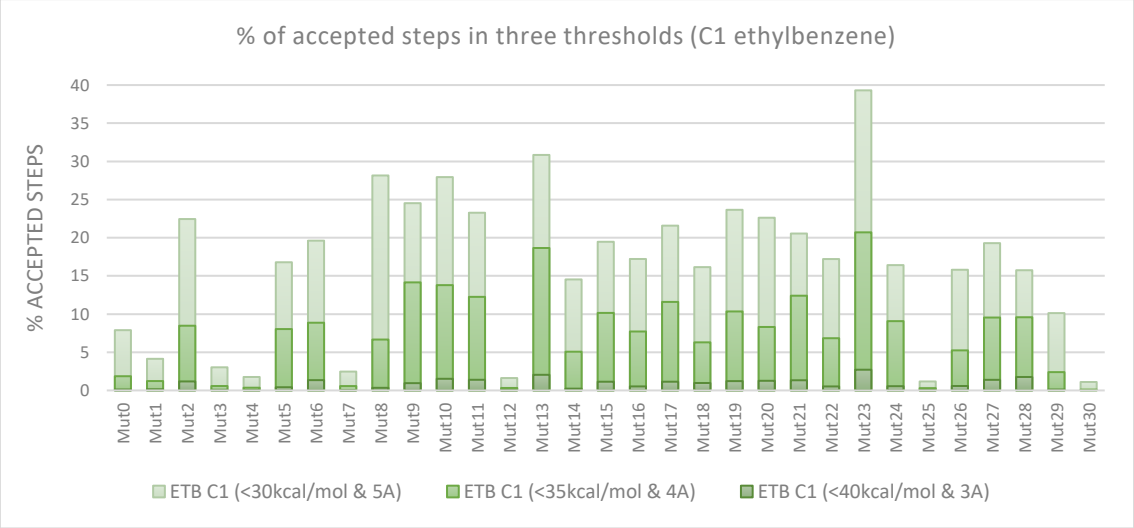


Figure 34. Thresholds of the accepted steps with the C1 of ethylbenzene. All mutants are represented.

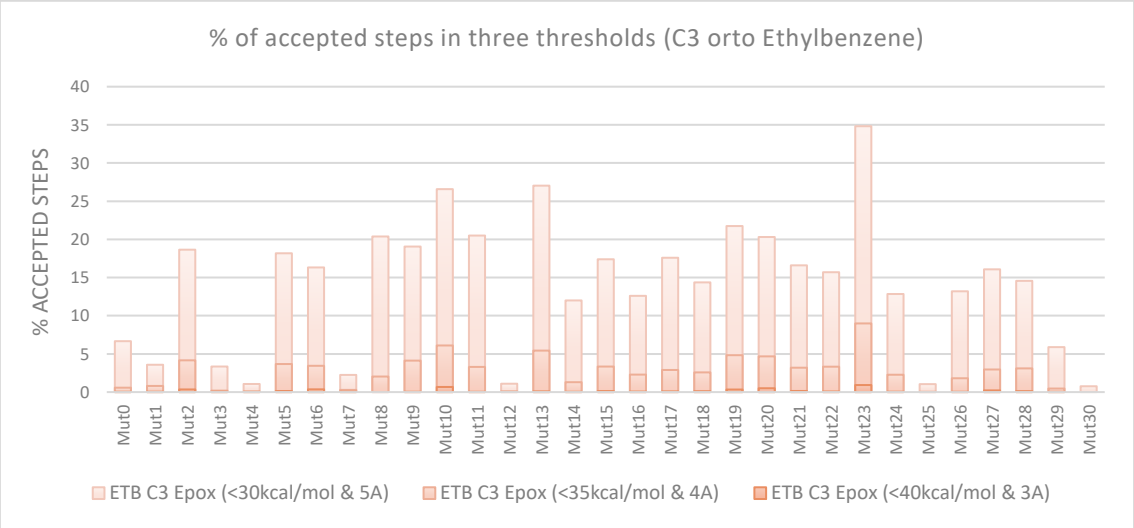


Figure 35. Thresholds of the accepted steps with the C3 of ethylbenzene. All mutants are represented.

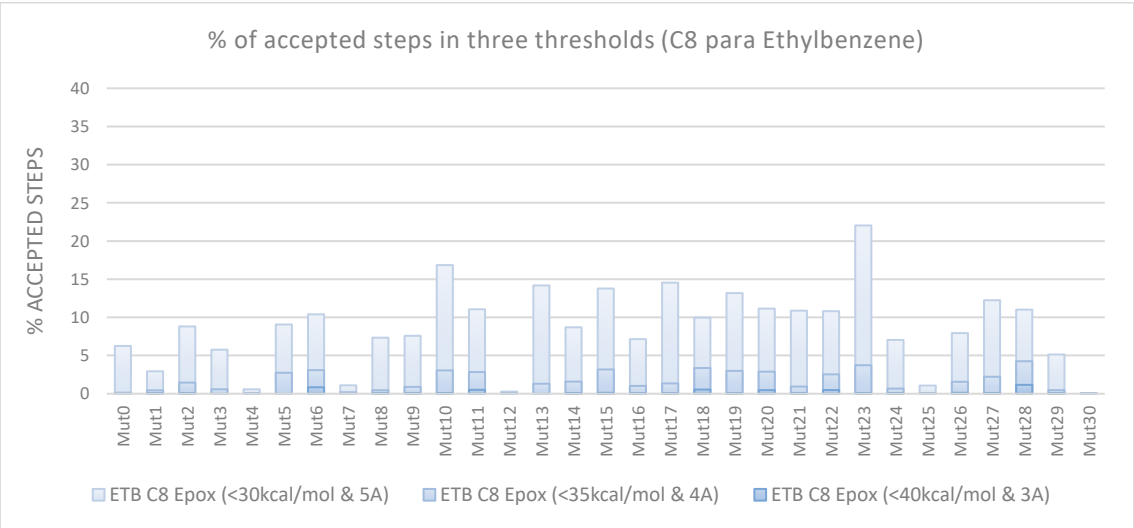


Figure 36. Thresholds of the accepted steps with the C8 of ethylbenzene. All mutants are represented.

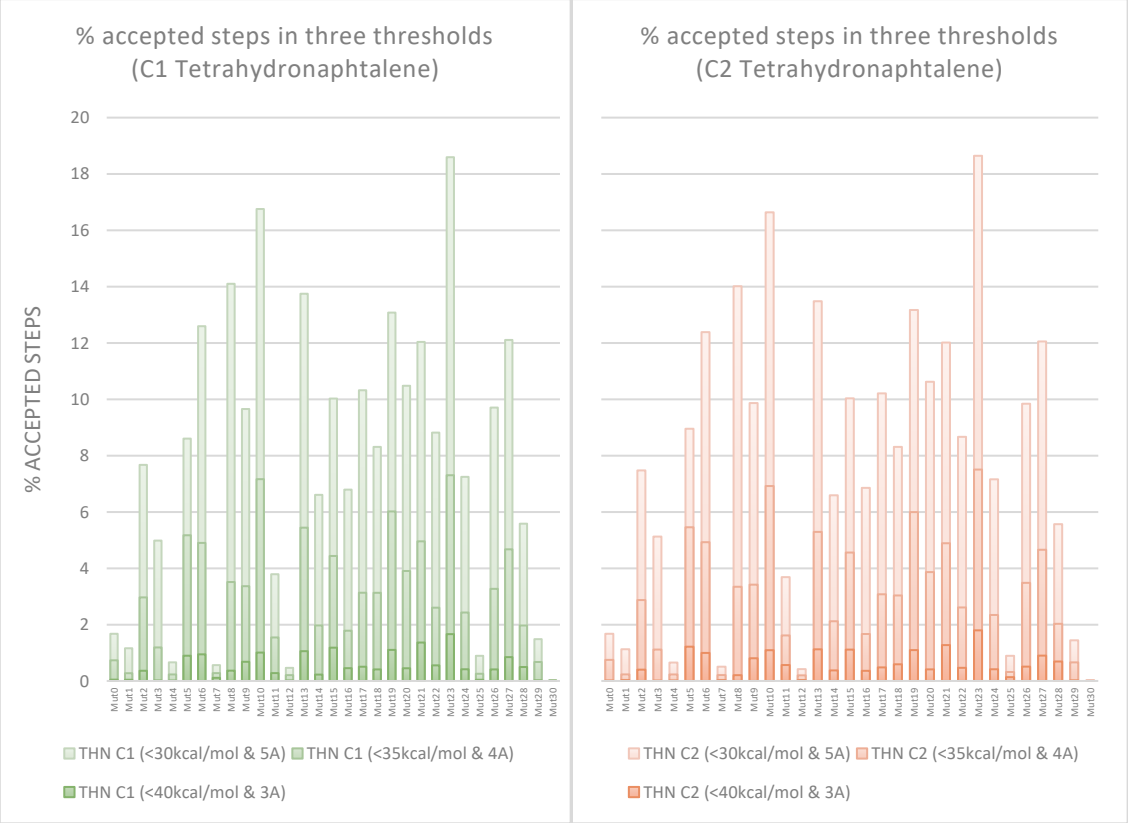


Figure 37 and Figure 38. Thresholds of the accepted steps with the carbon 1 and the carbon 4 of 1,2,3,4-tetrahydronaphtalene, respectively. All mutants are represented.

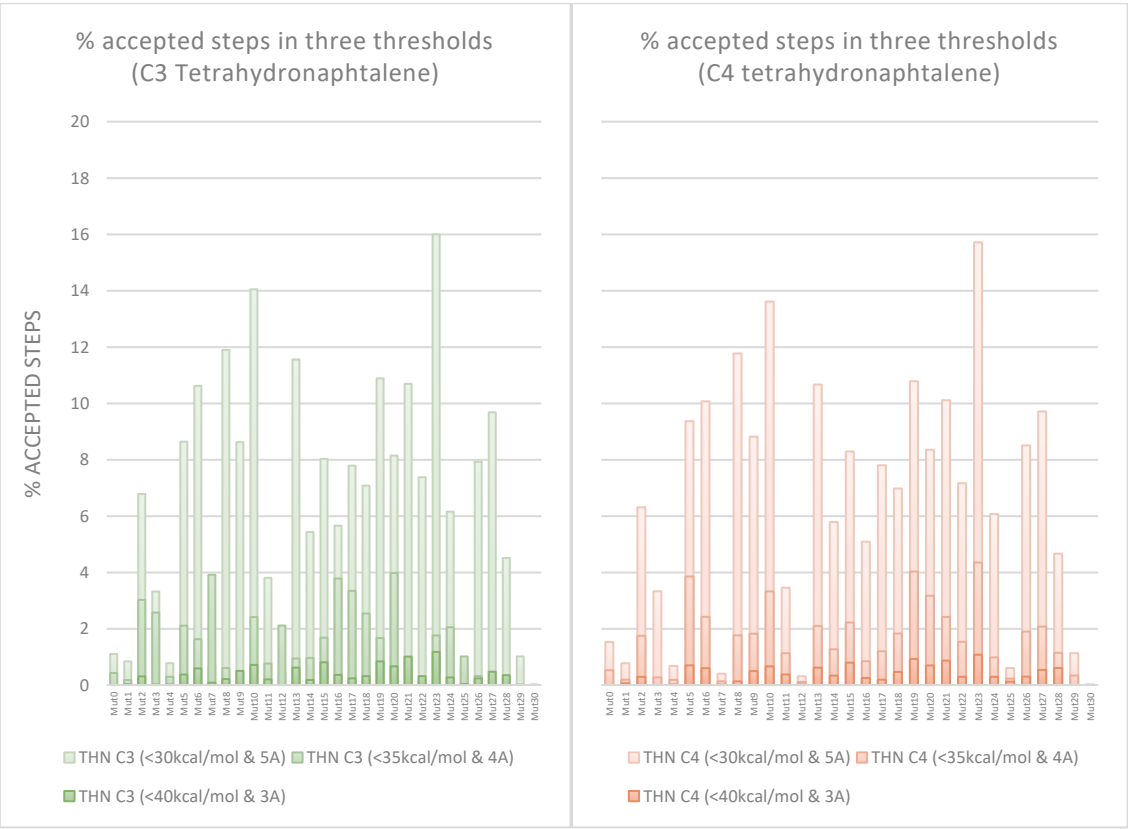


Figure 39 and Figure 40. Thresholds of the accepted steps with the carbon 2 and the carbon 3 of 1,2,3,4-tetrahydronaphtalene, respectively. All mutants are represented.

11.2. Residence time of all the mutants

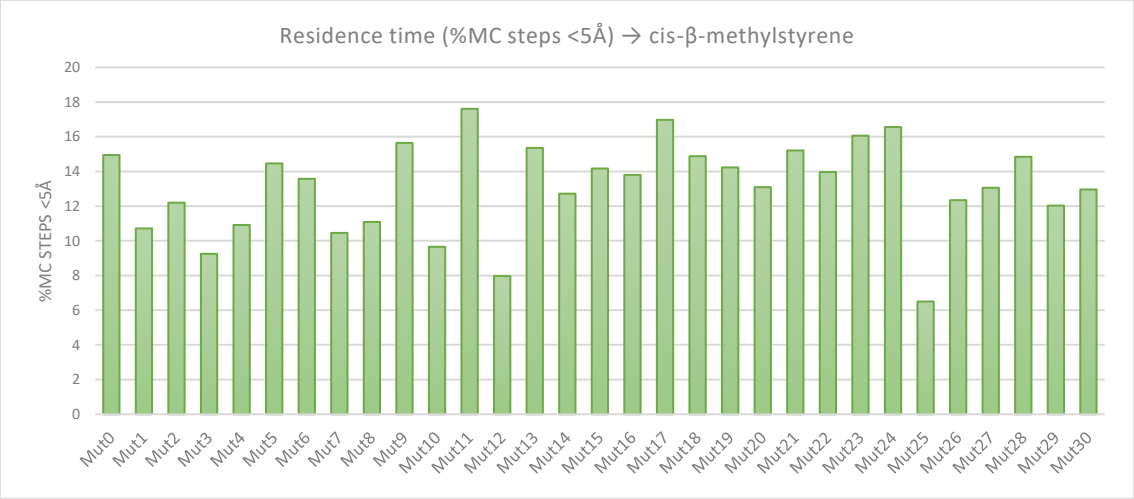


Figure 41. Residence time (% MC steps) under 5.0Å of every mutant. Distance between the Oxo and the middle of the double bond in cis-β-methyl-styrene.

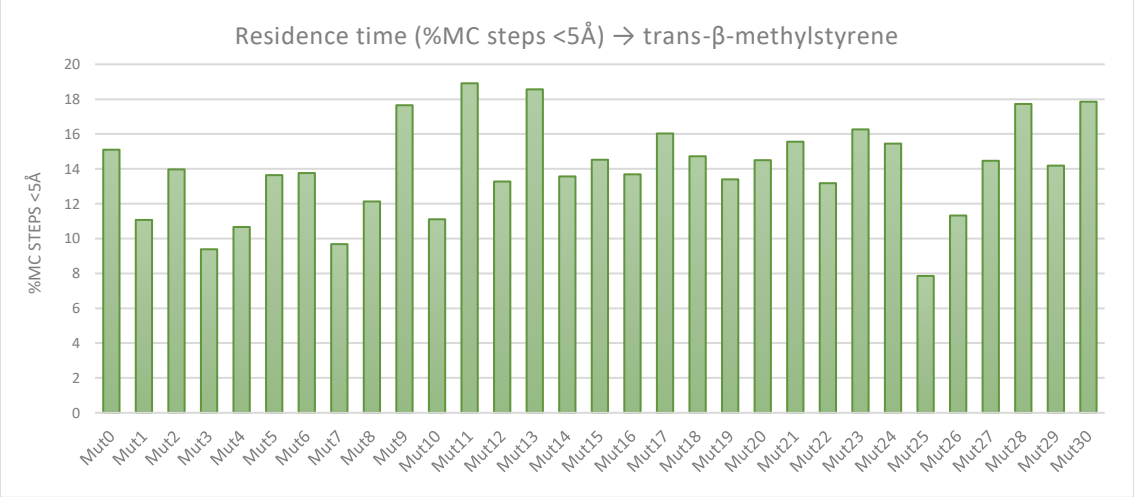


Figure 42. Residence time (% MC steps) under 5.0Å of every mutant. Distance between the Oxo and the middle of the double bond in trans-β-methyl-styrene.

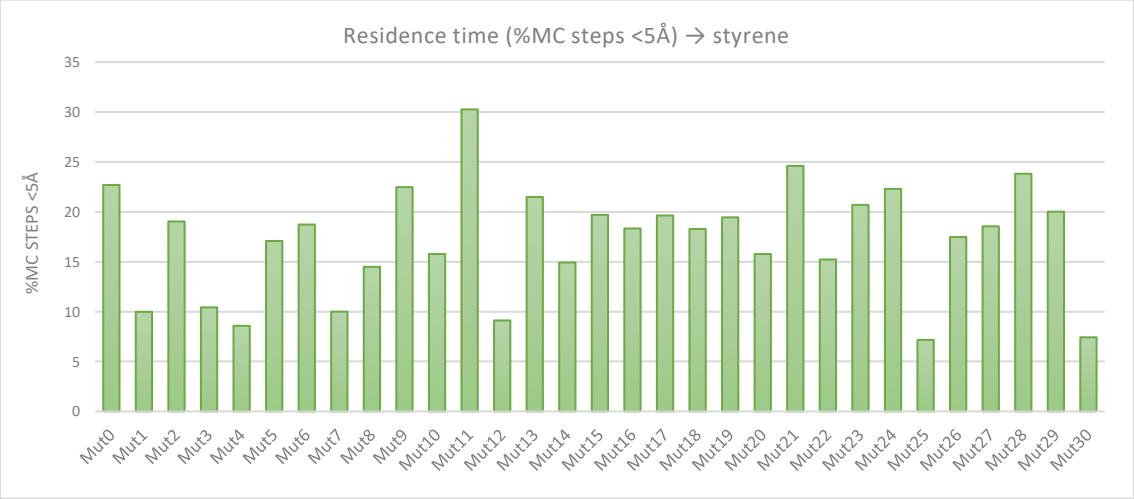


Figure 43. Residence time (% MC steps) under 5.0Å of every mutant. Distance between the Oxo and the middle of the double bond in styrene.

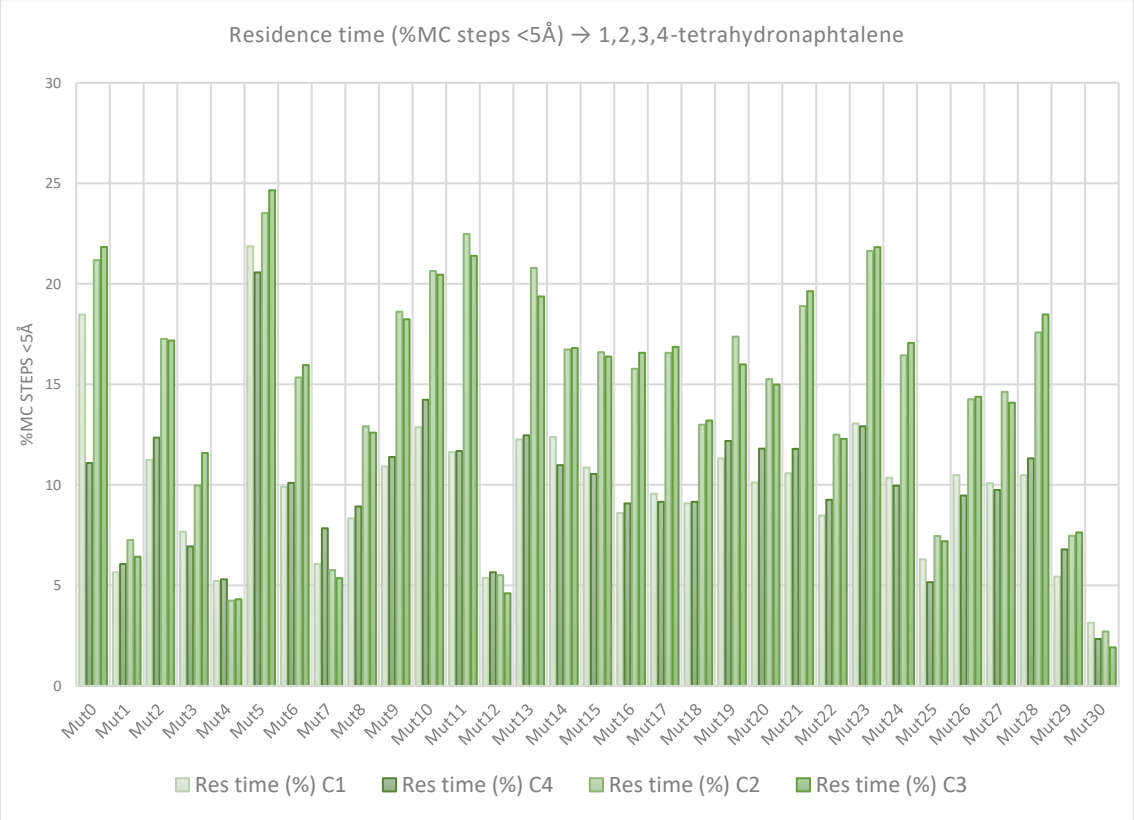


Figure 44. Residence time (% MC steps) under 5.0Å of every mutant. Distance between the Oxo and carbons 1, 2, 3, and 4 in 1,2,3,4-tetrahydronaphtalene.

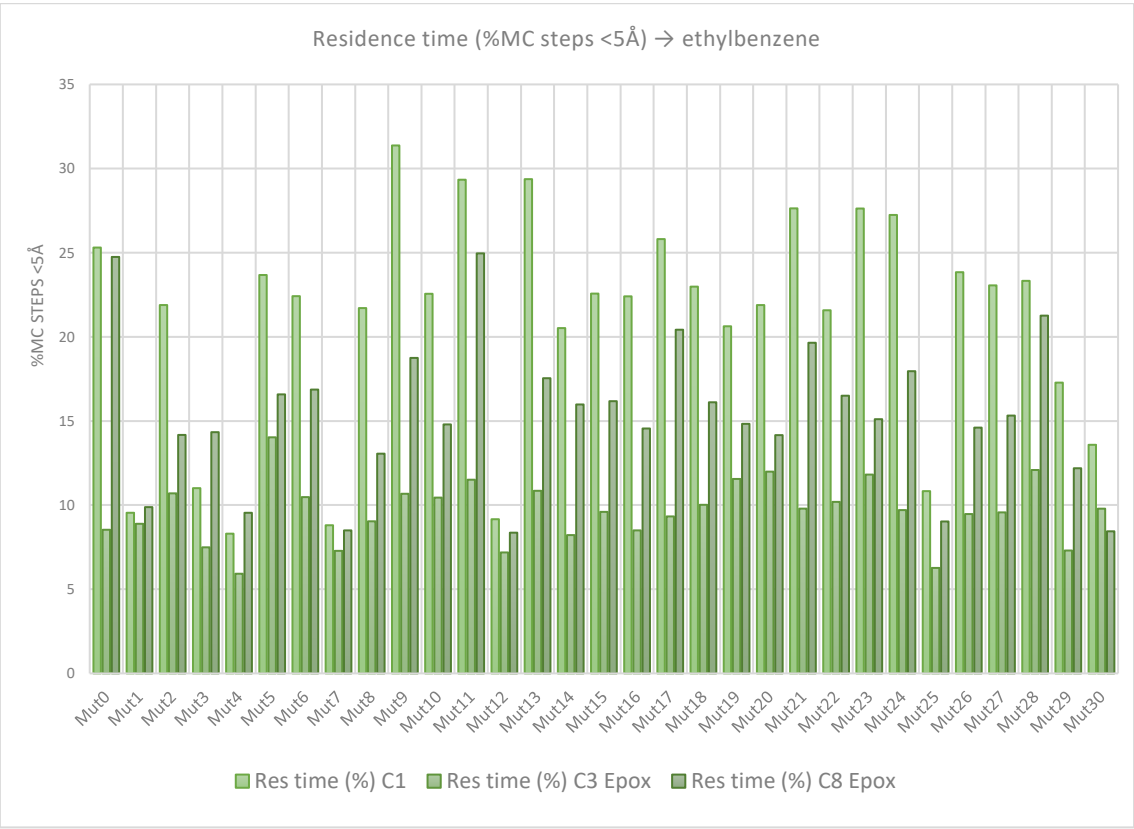


Figure 45. Residence time (% MC steps) under 5.0Å of every mutant. Distance between the Oxo and the carbon 1 and the middle of the double bond in carbons 3 and 8 in ethylbenzene.