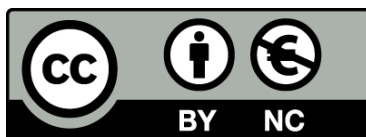UNIVERSITAT DE
BARCELONA

In silico molecular modelling and design
of heme-containing peroxidases
for industrial applications

Marina Cañellas Fontanilles
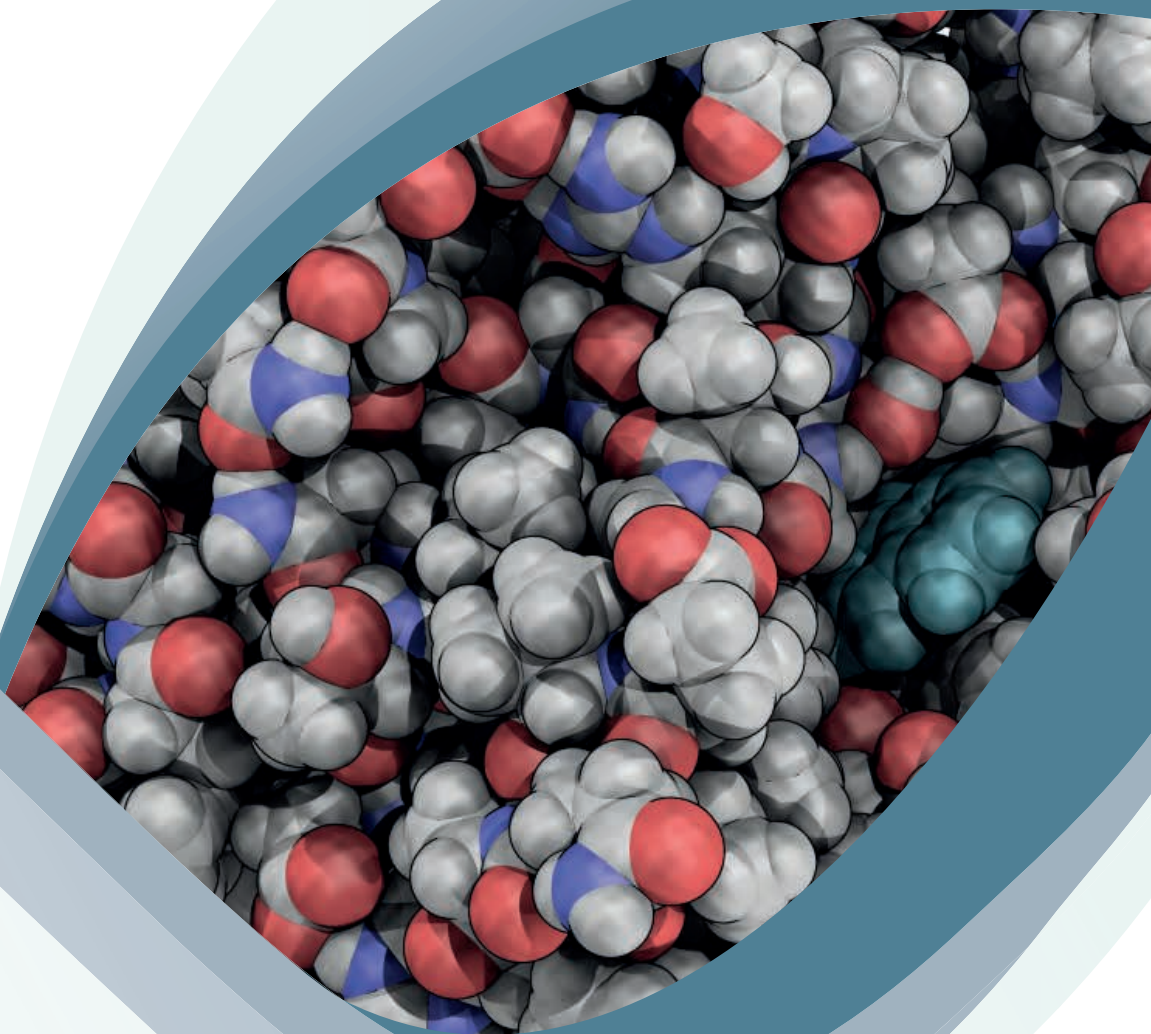
In silico molecular modelling and design
of heme-containing peroxidases for industrial applications
Marina Cañellas Fontanilles

UNIVERSITAT DE BARCELONA

Facultat de Farmàcia i Ciències de l'Alimentació

Programa de Doctorat en Biotecnologia

In silico molecular modelling and design of heme-containing peroxidases for industrial applications

Memòria presentada per Marina Cañellas Fontanilles per optar al títol de doctor per la Universitat de Barcelona

Dirigida per:

Dr. Victor Guallar Tasies   Dr. Maria Fátima Lucas

Tutora: Dr. Josefa Badia Palacín

Marina Cañellas Fontanilles
Barcelona, 2018

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

UNIVERSITAT DE BARCELONA

"Voici mon secret: **L'essentiel est invisible pour les yeux.***"*

"And now here is my secret: what is essential is invisible to the eye."

Antoine de Saint-Exupéry, *Le Petit Prince*

Table of Contents

ACKNOWLEDGEMENTS

To thank all the people that in one way or another have helped me in the realization of this thesis is not a trivial task. As an enzyme, this thesis is the result of a great number of small contributions that have given it the shape it has now. It is not only the binding site that matters, but also the remaining elements. That is why I will take my time to acknowledge all the people that have helped me to improve this work or have made this four-year road easier (sometimes even without noticing). Without all these tiny participations this thesis would not have been the same. This section has been written in distinct languages, in an attempt to exemplify that science does not care about frontiers.

Així doncs, comencem pel principi, per qui em va obrir les portes del seu grup permetent-me realitzar aquesta tesi: en Victor Guallar i la Fátima Lucas. Moltes gràcies Victor per donar-me aquesta increïble oportunitat que m'ha fet créixer tant en l'àmbit professional com en el personal. Gràcies per guiar aquest treball cap a la feina ben feta, per tots els teus suggeriments, idees i correccions. Un especial agradecimiento a Fati. Gracias por ser mi guía durante todo este viaje, por toda la ayuda que me has brindado, por tus recomendaciones y por tus "¡sí se puede!". Ha sido un autentico placer trabajar contigo!

En segon lloc, gràcies a totes les persones que en algun moment o altre han format part del meu grup al BSC. Gràcies per tots els bons moments (i pels que encara queden!), per la ciència compartida i la feina feta, perquè amb molts de vosaltres m'he guanyat una amistat i perquè un bon ambient a la feina ho fa tot més fàcil! Especialment, mil gràcies a en Ferran, a l'Emanuele i a en Gerard, heu fet que aquest camí sigui molt més planer! També he d'agrair a aquelles persones del departament de Life Sciences que, tot i no ser implícitament del meu grup, també han sigut peces clau en aquest procés. Gràcies per les converses, viatges, i bons moments en general, ha sigut un autèntic plaer!

También tengo que agradecer a todos los colaboradores del proyecto INDOX con quien he tenido el placer de trabajar: los miembros del grupo de Ángel Martínez en el CSIC-CIB, de Miguel Alcalde en el CSIC-ICP y de Ana Guitérrez en el CSIC-IRNAS. En especial, gracias a Patricia Molina, Patricia Gómez, Dolores Linde y Esteban Babot por el increíble trabajo experimental.

A la meva tutora de la tesi a la UB, Josefa Badía, i als membres de la SED de la facultat de farmàcia per tota l'ajuda que m'han proporcionat durant aquests anys. També vull agrair als responsables del projecte NOVADOMUS a la UB per a

donar-me la increïble oportunitat de poder fer una estada d'11 mesos a la Universitat d'Ottawa.

A special section goes for all the people that I had the pleasure to meet and to work with while doing my Novadomus stay at Prof. Roberto Chica Lab, at the University of Ottawa. Thanks to Roberto and to all the members of The Chica Lab for treating me as one more in your group. Although the work I performed there is not directly included in this thesis, having the opportunity of learning about Computational Protein Design techniques will be of crucial importance for my professional career in the near future. Dejando de lado el ámbito profesional, la estancia en Canadá me aportó muchas cosas (y personas) positivas, que aún sin estar implícitamente relacionadas con este trabajo, han contribuido a mi felicidad. Gracias a todas esas personas, que aún estando muy lejos, me he traído desde Canadá. En especial, gracias a Javi: quien tiene un amigo, tiene un tesoro.

De Canadá también te traje a ti. Gracias Juancho por haber compartido el final (¡y la peor parte!) de este largo camino conmigo, por tu apoyo y por tu paciencia. Gracias por enseñarme el verdadero significado de la palabra amar.

També vull donar les gràcies a la meva família. Gràcies mare, pare, Albert i Joan pel vostre suport incondicional. Encara que no us ho demostri massa sovint us estic infinitament agraïda. Muchas gracias a ti también Itziar para darle el toque artístico a esta tesis, ¡la portada es una pasada! I per acabar, només em queda agrair tot el suport i bons moments a la família que s'escull, als amics de Torredembarra i als de la carrera!

LIST OF PUBLICATIONS

This thesis is based on the following articles, which are referred to in the text by their Roman numerals.

I. Cañellas M., Linde D., Pogni R., Lucas F., Guallar V., Barratto M.C., Sinicropi A., Sáez-Jiménez V., Coscolín C., Romero A., Medrano F.J., Ruiz-Dueñas F.J. and Martínez A.T. (2015) Catalytic surface radical in dye-decolorizing peroxidase: a computational, spectroscopic and site-directed mutagenesis study. *Biochemical Journal.* 466: 253-262. **doi:** 10.1042/BJ/10141211

II. Babot E.D., del Río J.C., Cañellas M., Sancho F., Lucas F., Guallar V., Kalum L., Lund H., Gröbe G., Scheibner K., Ullrich R., Hofrichter M., Martínez A.T. and Gutiérrez A. (2015) Steroid Hydroxylation by Basidiomycete Peroxygenase: a Combined Experimental and Computational Study. *Applied and Environmental Microbiology.* 81(12): 4130-4142. **doi:** 10.1128/AEM.00660-15

III. Cañellas M., Lucas F., Babot D.E., del Río J.C., Kalum L., Ullrich R., Hofrichter M., Guallar V., Martínez A.T. and Gutiérrez A. (2016) Molecular determinants for selective C_{25}-hydroxylation of vitamins D_2 and D_3 by fungal peroxygenases. *Catalysis Science & Technology.* 6:288-295. **doi**: 10.1039/C5CY00427F

IV. Cañellas M., Linde D., Coscolín C., Davó-Siguero I., Romero A., Lucas F., Ruiz-Dueñas F., Guallar V. and Martínez A.T. (2016) Asymmetric sulfoxidation by engineering the heme pocket of a dye-decolorizing peroxidase. *Catalysis Science & Technology.* **6**: 6277-6285. **doi**: 10.1039/C6CY00539J

V. Molina-Espeja P., Cañellas M., Plou F.J., Hofrichter M., Lucas F., Guallar V. and Alcalde M. (2016) Synthesis of 1-Naphthol by a natural peroxygenase engineered by directed evolution. *ChemBioChem.* **17**: 341-349**. doi**: 10.1002/cbic.201500493

VI. Gomez de Santos P., Cañellas M., Tieves F., Younes F.H.H., Molina-Espeja P., Hofrichter M., Hollmann F., Guallar V. and Alcalde M. (2018) Selective synthesis of the human drug metabolite 5'-hydroxypropranolol by an evolved self-sufficient peroxygenase. *ACS Catalysis.* **8**: 4789-4799. **doi:** 10.1021/acscatal.8b01004

VII.   <u>Cañellas M.,</u> Sancho F., Guallar V. and Lucas F. (2018) Towards the efficient production of FDCA: Molecular Modelling and *In Silico* Design of *Aae*UPO, HMFO and AAO enzymes. (*Draft manuscript*)

OUTLINE OF THE THESIS

The present thesis has been drawn up as a compendium of publications and draft manuscripts. It has been divided into 6 different chapters, including a general *introduction* (Chapter 1), the *objectives* of the thesis (Chapter 2), the *results*, where *Articles I-VII* are presented (Chapter 3), a *summary of the results and discussion* of the work showed in Chapter 3 (Chapter 4), the overall *conclusions* (Chapter 5) and the general *references*, excluding those included in the here presented publications (Chapter 6).

Since this thesis is presented as a compendium of publications, I have classified each work (*Articles I-VII*) accordingly to the question that aims to be answered rather than the enzyme system that is studied. In this way, each section of Chapters 2-5, will include or will be related with a number of publications (either published or in process to be published) attempting to achieve the following objectives: i) Reaction mechanisms description, ii) Enhanced variant rationalization, and iii) Computational enzyme redesign.

Moreover, it is important to note that the realization of *Articles I-VII* has been only possible through the collaboration of a number of scientists and research groups very diverse in their expertise areas. The entire articles will be presented in Chapter 3 of this thesis. However, only the computational section of these publications has been performed by myself. Accordingly, and to avoid false assumptions of authorship, Chapters 4-5 include only the results, discussion, and conclusions obtained during the development of this thesis.

ABBREVIATIONS

5'-OHP	5'-hydroxypropranolol
*Aae*UPO	*Agrocybe aegerita* unspecific peroxygenase
AAO	Aryl-alcohol oxidase
*Aau*DyP	*Auricularia auricular-judae* dye-decolorizing peroxidase
ABTS	2 2'-azino-bis(3-ethylbenzothiazoline-6-sulfonate)
ANM	Anisotropic Network Model
BO	Born-Oppenheimer
*Cci*UPO	*Coprinopsis cinerea* unspecific peroxygenase
CPO	Chloroperoxidase
DE	Directed Evolution
DFT	Density Functional Theory
DMP	2,6-dimethoxyphenol
DNA	Deoxyribonucleic acid
DyP	Dye-decolorizing peroxidase
ET	Electron Transfer
FDCA	2,5-furandicarboxylic acid
FF	Force Field
FFCA	2,5-formylfurancarboxylic acid
HF	Hartree-Fock
HK	Hohenberg-Kohn
HMF	5-hydroxymethylfurfural
HMFCA	2,5-hydroxymethylfurancarboxylic acid
HMFO	Hydroxymethylfurfural oxidase
HTP	Heme-thiolate peroxidases
IFD	Induced Fit Docking
k_{cat}	Catalytic constant

k_{cat}/K_m	Catalytic efficiency
K_m	Michaelis-Menten constant
KS	Kohn-Sham
LCAO	Linear Combination of Atomic Orbitals
LRET	Long-Range Electron Transfer
MC	Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics
MPS	Methyl-phenyl sulfide
MTS	Methyl-*p*-tolyl sulfide
OPLS	Optimized Potential for Liquid Simulations
P450	Cytochrome P450
PBC	Periodic Boundary Conditions
PDB	Protein Data Bank
PELE	Protein Energy Landscape Exploration
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
RB19	Reactive Blue 19
TISE	Time-Independent Schrödinger Equation
TON	Turnover number
TS	Transition State
UPO	Unspecific peroxygenase

Introduction

1

"The greatest enemy of knowledge is not ignorance. It is the illusion of knowledge."

Stephen Hawking

1.1 ENZYMES

Life, as we know it, is based on a series of well-organized chemical reactions. Although many of these reactions occur spontaneously, their transformation rates are too low to be actually capable of sustain life [1, 2]. In this way, nature, side by side with the evolution machinery, has encountered the perfect solution: enzymes. Enzymes are catalytically active proteins that speed up the rates of chemical reactions, hence enabling essential transformations for the life and function of all sort of organisms, from viruses to the most evolved forms of life [3]. Interestingly, the potential of such biological catalysts can overcome the physical boundaries of living organisms, and many enzymes still preserve their catalytic activity once extracted from the host. Because of this, humankind did not take a long time to become aware of the potential underneath enzymes and to use them for their benefit.

1.1.1 Brief history of enzymes

The earlier shreds of evidence of enzyme's potential exploitation by humans are dated from the ancient era. Texts from the early civilizations of Babylon, Greece, Egypt, China, and India refer to the use of microorganisms as enzyme sources for the production of cheeses, bread and alcoholic beverages, and also for the meat tenderization and the manufacture of goods such as indigo, linen, and leather [1, 4-6]. However, while Ancient people blindly performed those processes relying essentially on empirical evidence and folklore, they were not conscious that such transformations were caused by enzymes or about the chemical principles beneath them.

It was not until the early nineteenth century that scientists' attention began to focus on studying enzymes more systematically. As a matter of fact, the first reference we have of an article describing a soluble extract with enzymatic activity is dated from 1833 [7, 8], in which Payen and Persoz described the preparation of diastase, a soluble extract that cleaved starch. By the second half of the nineteenth century, several studies about fermentation processes and catalytic activities were performed. However, there was considerable debate about the nature of the enzymatic activity: Was it solely originated by the intact cell, as postulated by Luis Pasteur and other scientists, or was it due to the action of some molecular components within the cell? Eduard Buchner answered such question with the discovery of the cell-free fermentation, which led him to earn the Nobel Prize in Chemistry in 1897 [7, 9].

After Buchner's pioneering work, the next step that followed was to determine the true biochemical nature of enzymes that was established by James B. Sumner in 1926. He crystallized the first enzyme, the urease from jack bean, and revealed the true nature of enzymes as proteins. Such achievement led him to earn the Nobel Prize in Chemistry in 1947 [4, 7, 9]. At that point, with Buchner and Sumner achievements, along with x-ray crystallography, the era of structural biology began, providing insights to understand the mechanism of action of enzymes at the molecular level and enabling the use of biocatalysts for an inconceivable number of applications. **Fig. 1.1** shows some of the milestones achieved from the beginning of the nineteenth century regarding enzymes.
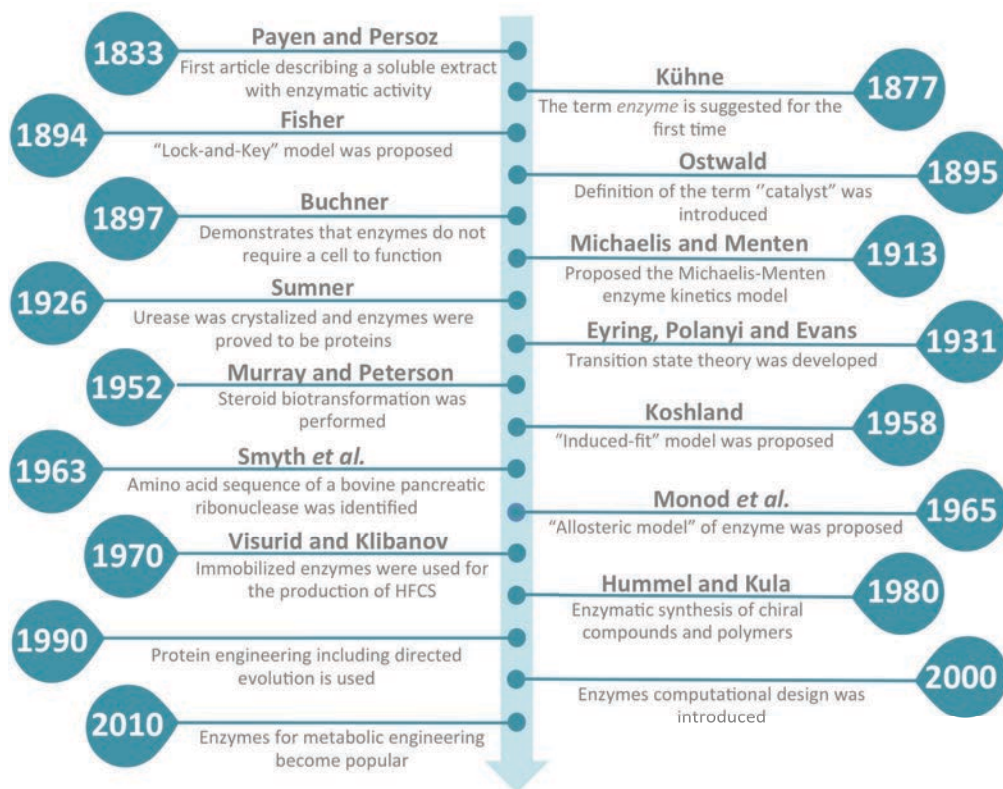
Figure 1.1. Timeline of important events and discoveries concerning enzymes from the early nineteenth century up to now [4, 7, 9-16].

1.1.2 Enzymes: the perfect catalysts

Although fundamental questions still remain unraveled regarding enzymes, all the knowledge accumulated up to now has given us a whole clear picture of the mechanisms of action and structure of such catalysts. Enzymes are life-essential naturally occurring or synthetic biological catalysts (also known as biocatalysts)

that speed up the reaction rates by mean of lowering the activation energy and without themselves undergoing any change [2, 17]. Living cells have evolved by adapting to two contradictory needs: Their cells have to be stable under most conditions, but yet cells must be dynamic entities. Enzymes enable this incongruity by maintaining and allowing such dynamic changes in a controlled and time-effective way, facilitating the essential molecule and protein modifications in the cell [7]. Thus, enzymes are indispensable to the maintenance and activity of life, being required in most of the reactions involved in metabolism, as well as deoxyribonucleic acid (DNA) manipulation, respiration, and digestion [18]. Interestingly, they can also be extracted from cells, allowing the use of their catalytic power to perform a wide variety of commercially important processes [17].

Natural amino acids
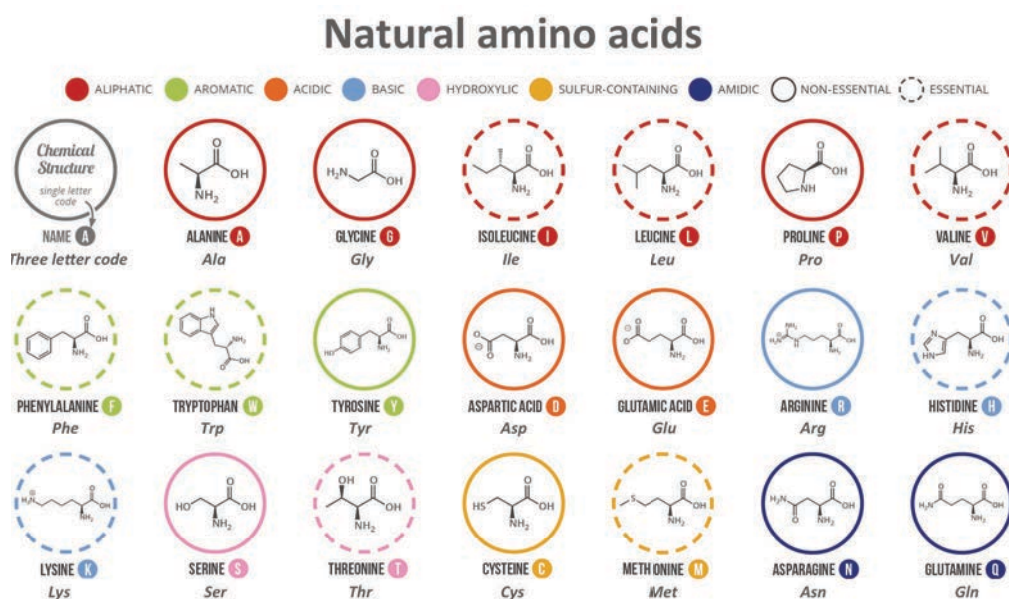
Figure 1.2. Chemical structure of the 20 natural amino acid, along with their names, three letter, and single letter codes. (Image modified from http://www.compoundchem.com/20-14/09/16/aminoacids/)

From a structural point of view, the grand majority of functional enzymes are globular proteins. However, it must be noted that a small subset of biocatalysts does not have a globular proteic nature. This is the case of ribozymes, ribonucleic acid (RNA) molecules that can exert catalytic effects, and 'abzymes', human-made antibodies that possess catalytic properties [17]. Like other proteins, enzymes are large macromolecules consisting of chains built up mainly by the 20 naturally occurring amino acids (**Fig. 1.2**), which are linked together by peptide bonds. Enzymes can range in size from less than 100 to more than 2000 residues and can be

arranged as a single polypeptide or as a subunit in a multiple polypeptide chain enzyme. Each residue in a polypeptide chain is constituted by an alpha carbon atom linked to an amino group, a carboxyl group, a hydrogen atom, and a variable side-chain that will determine the chemical nature of each residue. The amino acid sequence constituting the polypeptide chains is unique for each enzyme and will determine the short-range structural patterns (the secondary structure) such as loops, helices, and sheets, which will be spaciously arranged to form the three-dimensional structure of the protein. This configuration, which is essential for the enzyme's activity, is stabilized by numerous interactions between the amino acids in different parts of the peptidic chains (**Fig. 1.3**) [17, 19].

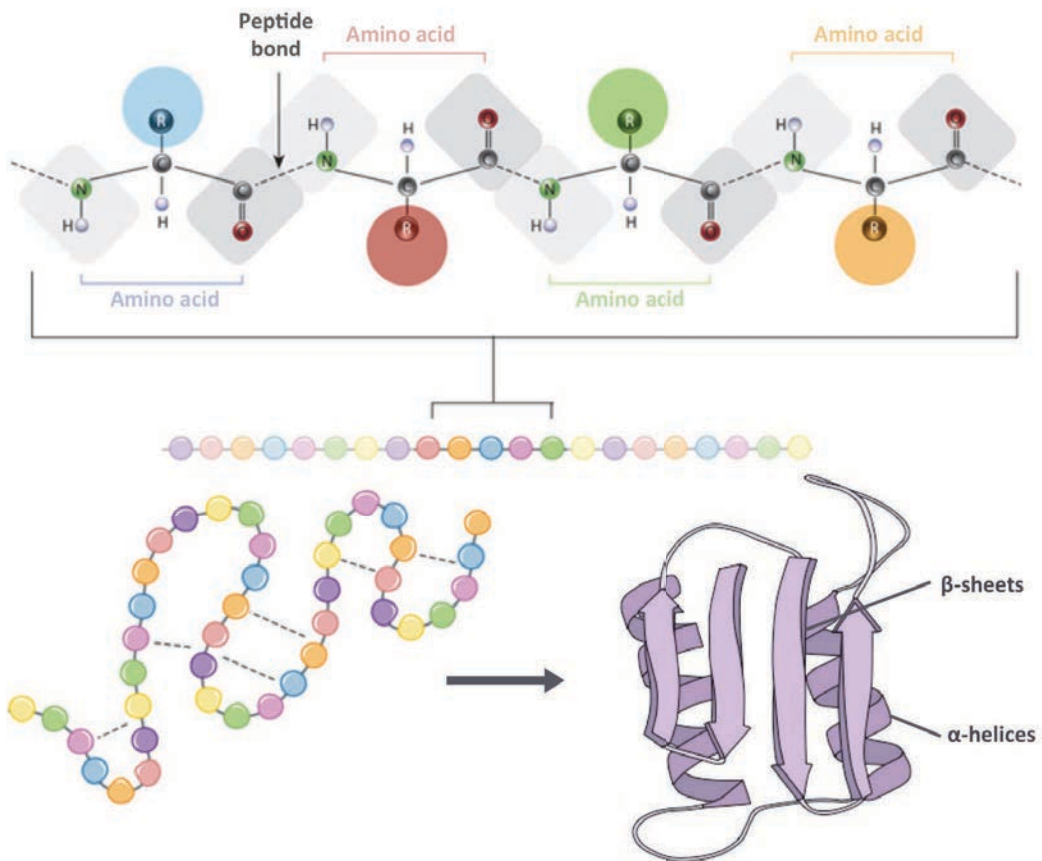**Figure 1.3.** Structural organization of globular proteins. Different amino acid residues with unique side-chains are connected together by peptide bonds to constitute a polypeptide. The peptidic chain will then fold into a specific conformation depending on the interactions between amino acid residues [19]. (Image modified from O´Connor *et al.*, 2010)

From a mechanistic perspective, only a small part of a protein is directly involved in the catalytic event: the active site. The enzyme's active site has a unique 3D architecture, with particular electrostatic charge patterns and hydrophobicity, and is the principal responsible for the activity and selectivity of the biocatalyst. Among the active site residues, only a few (2-4) are directly involved in catalysis; the remaining residues ensure a proper substrate positioning to facilitate the reaction. Besides, numerous enzymes require additional non-proteic chemical compounds or metallic ions, named cofactors, for an efficient catalytic activity. Cofactors usually are placed on the active site of the enzyme or nearby it.

Because of their nature and structure complexity, enzymes hold several properties that make them very appealing for industrial applications. Hence, biocatalysts are seen as a possible alternative to conventional chemical catalysts, offering multiple advantages over them. Due to the high level of complexity and structuralization of their active site, enzymes can exhibit a high degree of enantio-, chemo-, regio- and choro-selectivity [20]. Moreover, enzymes' catalytic efficiency is surprisingly high, showing turnover numbers (TONs) of $>10^5$ s^{-1} compared with the values of 10^{-2} s^{-1} usually observed by heterogeneous and homogeneous chemical catalysts [21, 22]. Since enzymes have been naturally evolved to function in physiological conditions, they usually work under mild conditions of temperature, pressure, and pH, and in an aqueous environment, conditions that decrease the energy requirements and consequently the capital costs of reactions in an industrial scale. In addition, enzymes are inherently non-hazardous, reusable and biodegradable catalysts, making them ideal environmentally friendly reagents [20, 23, 24].

Paradoxically, the advantages that suppose the use of biocatalysts can sometimes also be a drawback. Nature has tailor-made enzymes to become proficient catalysts in a physiological environment. Because of this, naturally occurring catalysts can be inhibited by the conditions required in some industrial processes, such as high temperatures, extreme pH conditions, non-aqueous solvents or high concentrations of end products [4, 23, 25]. Furthermore, enzymes tend to be captious with the substrates they are willing to accept, and their affinities for different substrates can vary a lot. Industrial processes need catalysts capable of working with various substrates at high efficiencies, in particular conditions, sometimes with unnatural substrates, and additionally, these biocatalysts must be cost-effective [26]. In this way, native enzymes can be tailored towards optimal physicochemical and functional properties with the use of enzyme engineering strategies [25, 26].

1.1.3 Use of enzymes for industrial & technological applications

As previously mentioned, since ancient times, naturally occurring enzymes have been used for the production of food and other commodities. However, it was not until 40 years ago that the enzyme industry began to experience a fast growth due to the development of modern biotechnology, including recombinant DNA, high-throughput screening, and protein engineering technologies [23]. Because of that, their applications have gained extensive use in the textile, grain processing, animal feed, detergent, pulp and paper, fuel ethanol, and chemical industries. **Table 1.1** shows some of the industrial enzyme applications based on different industrial sectors [23, 27]. From an economic perspective, the world market for industrial enzymes approached $5 billion/year in 2013 and was growing at a rate of 8% per annum, expecting global sales of $7.1 billion by 2018 [28]. Two companies dominate enzymes' global production: the Danish company Novozymes, with a market share of 47%; and the U.S. company Dupont (that has recently acquired Genencor), with 21% [17].

With thousands of different enzymes being isolated and characterized, one would think that their industrial applications are innumerable. However, less than 100 enzyme types have ever been used in industrial applications, and only around 40 are produced on a truly industrial scale (production ranging from multiple kilograms to tons per annum) [17, 28]. As already mentioned, native enzymes usage is still limited in an industrial context mainly due to problems with substrate selectivity and efficiency, and long-term stability under specific process conditions. Some improvements in process efficiency can be achieved by adapting the chemical manufacturing process to the sensitivities of the biocatalysts. The other alternative is to use protein engineering methodologies to generate new enzymes with enhanced properties concerning stability, selectivity, and efficiency, followed by immobilization to establish more robust processes [29]. In these lines, protein engineering is of particular interest for both academia and industry, and lots of efforts are being made to study and improve natural existing enzymes for their potential use in industrial processes.

Table 1.1. Enzyme applications based on five different fields: Technical applications, food processing, animal feed industry, organic synthesis industry and cosmetic industry [23, 27].

Application fields	Enzyme	Applications
Technical applications — Pulp and paper industry	Amylases	Reducing viscosity in starch coatings
	Cellulases	Making fibers flexible
Textile industry	Laccases	Decolorization and detoxification of effluents from textile
	Peroxidases	Dye excess removal
Laundry detergents	Proteases	Protein-based stain removal
	Amylases	Removal of resistant starch residues
Biorefinery	Cellulases and hemicellulases	Degrading lignocellulosic materials
	Lipases	Biodiesel production
Food processing — Dairy industry	β-galactosidases	Avoiding lactose intolerance
	Chymosin, lipases and lysozymes	Cheese manufacturing
Starch industry	α-amilases and amyloglucosidases	Conversion of starch to glucose syrup
	Cyclodextrin glycosyltransferases	Cyclodextrins production
Baking industry	α-amilases	Controlling the structure of bread
	Glucose oxidases	Dough strengthening
Juice industry	Pectinases	Increasing juice production yield
	Laccases	Phenol derivatives removal
Brewing industry	Pullulanases	Securing maximum wort fermentability
	Acetolactate decarboxylases	Improving beer taste
Fat and oil industry	Lipases	Flavour production
	Phospholipases	Lyso-lesithin production
Animal feed industry	Phytases	Releasing phosphorus and increasing availability of cations
	Xylanases	Degrading fiber in viscous diets and increasing nutritive value of feed
Organic synthesis industry	Monooxygenases	Synthesis of optically pure epoxides
	Laccases	Oxidation of phenol derivatives
Cosmetic industry	Oxidases and peroxidases	Hair dyeing
	Papain, bromolain and subtilisin	Giving gentle peeling effects in skin care

1.2 SYSTEMS OF STUDY: HEME-CONTAINING PEROXIDASES

Peroxidases (EC 1.11.1.x) are secreted, microsomal or cytosolic enzymes distributed in all kingdoms of life that utilize hydrogen peroxide (H_2O_2) or organic hydroperoxides as co-substrates [30]. Most of them are heme peroxidases that use heme *b* (iron-protoporphyrin IX, structure shown in **Fig. 1.4A**) or post-translationally modified heme as redox cofactor to catalyze the hydrogen peroxide-mediated oxidations of a myriad of substrates, including aromatic molecules, cations, anions, or even proteins [31, 32]. Most heme peroxidases catalyze the one-electron oxidation of phenols, aromatic amines and carbon atoms (**Fig. 1.4B**, *reaction 1*), although some show also two-electron oxidation activity over halide and pseudohalide ions (**Fig. 1.4B**, *reaction 2*). Besides from the typical peroxidative activity, a very few heme peroxidases, such as chloroperoxidase (CPO, EC 1.11.1.10), also show catalase activity (**Fig. 1.4B**, *reaction 3*) [33]. Moreover, a specific superfamily of heme peroxidases, which share the binding site reorganization of cytochrome P450 (P450), also have monooxygenases-like peroxygenase activity (**Fig. 1.4B**, *reaction 4*), being able to catalyze selective oxygen transfers into organic molecules [31, 34]. Because of their catalytic versatility and enzymatic stability, there is a particular interest in the use of heme peroxidases for industrial applications, which is reflected in the increasing number of patents that have been published concerning peroxidases over the last years [30].

A **B**

$$H_2O_2 + 2AH_2 \rightarrow 2H_2O + 2 \cdot AH \quad (Reaction\ 1)$$

$$H_2O_2 + X^- + H^+ \rightarrow H_2O + HOX \quad (Reaction\ 2)$$

$$H_2O_2 + H_2O_2 \rightarrow 2H_2O + O_2 \quad (Reaction\ 3)$$

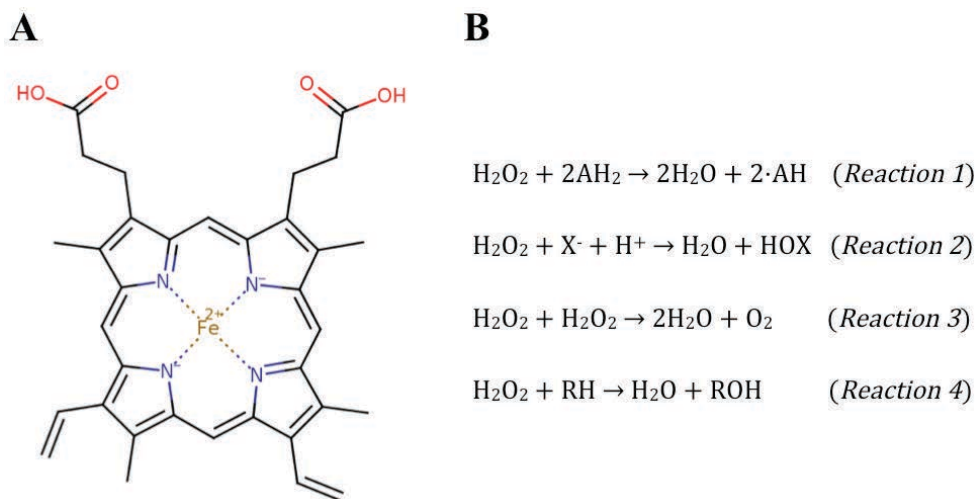$$H_2O_2 + RH \rightarrow H_2O + ROH \quad (Reaction\ 4)$$

Figure 1.4. (A) Structure of the iron-protoporphyrin IX. (**B)** Heme peroxidases common reactions [31].

Heme peroxidases are functional, structural and phylogenetically very diverse, and because of this, their classification has been subjected to extensive

debate [30-32, 35, 36]. However, the most generally accepted classification, which considers structural and catalytic properties rather than the enzyme origin, divides heme peroxidases into four different superfamilies: peroxidase-catalase, peroxidase-cyclooxygenase, peroxidase-chlorite dismutase and peroxidase-peroxygenase superfamily (previously known as heme-thiolate peroxidases (HTP) or haloperoxidases) [31]. In particular, this thesis is centered on the study of two fungus heme peroxidases: *Auricularia-auricula judae* dye-decolorizing peroxidase (*Aau*DyP), from the peroxidase-chlorite dismutase superfamily, and an unspecific peroxygenase (*Aae*UPO) from *Agrocybe aegerita* that has been recently classified into the peroxidase-peroxygenase superfamily. Besides, in one additional project, an unspecific peroxygenase from *Coprinopsis cinerea* (*Cci*UPO) is studied together with *Aae*UPO.
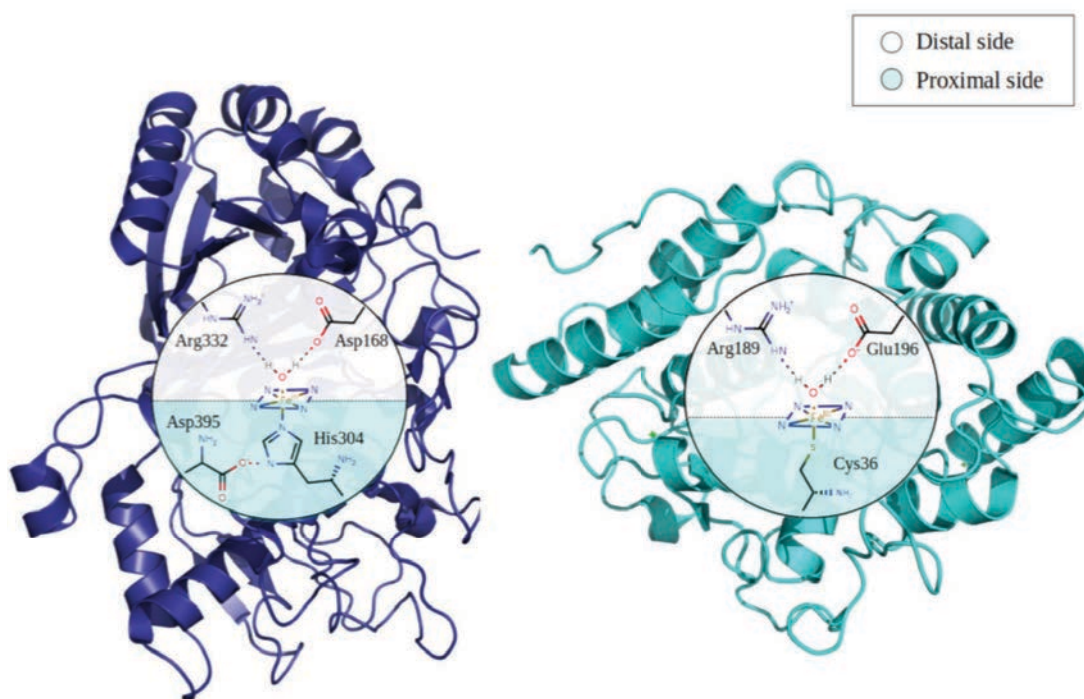
Figure 1.5. Structural model of two representative heme peroxidases from the peroxidase-chlorite dismutase and peroxidase-peroxygenase superfamilies. **Left panel** - Active site architecture of *Aau*DyP (PDB:4W7J) showing a schematic representation of the heme cofactor on its resting state (Fe^{3+} with an axial aqua ligand) along with its proximal (Asp395 and His304) and distal (Arg332 and Asp168) residues. **Right panel** - Active site architecture of the HTP *Aae*UPO (PDB:2YOR) showing the heme cofactor on its resting state along with its fifth ligand (Cys36) on the proximal side and distal (Arg189 and Glu196) residues.

As previously mentioned, heme peroxidases are structurally very diverse. Their size can range from tens to hundreds of amino acids, they can have one or multiple domains and variable numbers of sulfur bonds, sometimes the heme group can be post-translationally modified and covalently bound to the protein, they can bear ion-binding motifs, and their overall globular representative folds varies a lot [31]. Despite their general structural differences, heme peroxidases share a very similar catalytic cycle and binding site organization, where the heme prosthetic group is included in a hydrophobic pocket and surrounded by conserved residues that modulate the features of the cofactor and the overall enzyme's reactivity. The heme cofactor consists of a protoporphyrin complex and an iron cation, Fe^{3+} in the resting state, bound at the center and coordinated with four nitrogen atoms of the planar porphyrin ring (**Fig. 1.4A**). A histidine epsilon nitrogen atom or a cysteine thiolate anion, in the case of peroxidases from the peroxidase-peroxygenase super-family, serves as the fifth ligand to the heme iron on the proximal side [31, 32, 37, 38]. In most heme peroxidases, this fifth ligand is hydrogen bonded to an aspartate or a glutamate residue that collaborates to the stabilization of the iron high oxidation states by the so-called push effect [39-41]. This is not the case of HTPs, where the ligated cysteine thiolate exerts the push effect itself [42, 43]. The distal heme side bears the acid-base pair, which is essential for catalysis. Most heme peroxidases use a histidine or a deprotonated aspartate or glutamate residue as an acid-base catalyst, which is usually paired with an arginine. Distal residues are involved in the heterolytic cleavage of the hydrogen peroxide during the enzyme activation, and they are essential for the formation and stabilization of high valent ferryl intermediates of the heme iron, such as compound I [37, 38, 44]. **Fig. 1.5** shows the overall three-dimensional structure of the heme peroxidases studied in this thesis, along with their respective binding sites.

All heme peroxidases share the one-electron oxidation (**Fig. 1.4**, *reaction 1*) mechanism. This net transformation, however, comprises a more complex cycle that is schematized in **Fig. 1.6** (purple-colored reaction), and exemplified with 1-naphthol oxidation. The first step of the oxidation involves the binding of the hydrogen peroxide to the sixth coordination position of the heme iron on its resting state (Fe^{3+} with an axial aqua ligand), with the concomitant displacement of a water molecule. Then, the distal acid-base residue abstracts a proton from the peroxide producing a ferric hydroperoxide intermediate [45], commonly known as compound 0, which is stabilized by the distal arginine. It is thought that the distal acid-base catalyst acts again by delivering the previously acquired proton to the terminal compound 0 oxygen, catalyzing the O-O bond cleavage that produces compound I and forming a water molecule. This mechanism is referred as the Poulos-Kraut mechanism [46]. Moreover, it has been hypothesized that in some heme

peroxidases a water molecule could assist in the proton transfer process [37]. Compound I consists of a ferryl species coupled with a radical cation located on the porphyrin framework. Then, a substrate molecule is oxidized by the π porphyrin radical forming compound II, which is regenerated into the heme resting state by a second substrate molecule, releasing water. Furthermore, it has been demonstrated that in the absence of an oxidizable substrate, the compound I radical cation can also be localized in a binding site or surface tyrosine or tryptophan residue. This protein radical can act as an intermediate to oxidize substrates through a long-range electron transfer mechanism (LRET) [34].
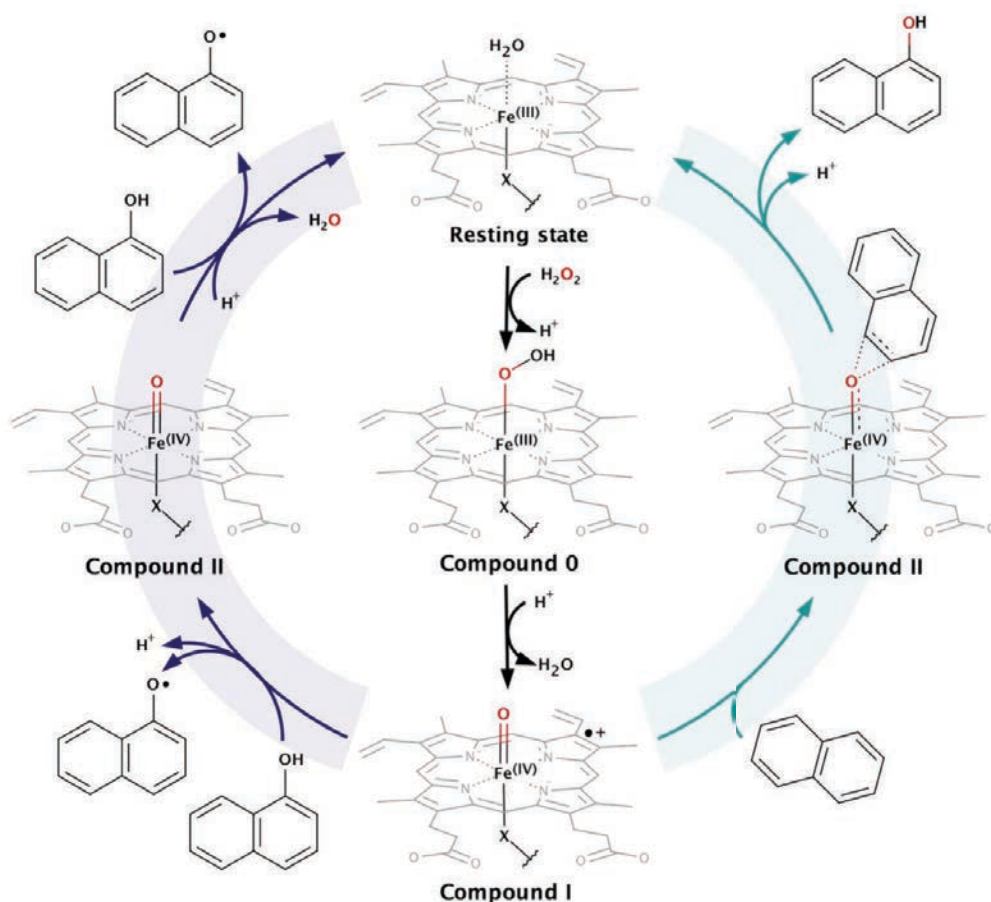
Figure 1.6. Catalytic cycle of heme peroxidases illustrated with the hydroxylation of naphthalene (cyan mechanism) and oxidation of 1-naphthol (purple mechanism). The hydrogen peroxide oxygen atom transferred in the peroxygenase mechanism is colored in red. The proximal residue, here represented with an X, can be a histidine or a cysteine residue (in the case of heme-thiolate enzymes). However, only thiolate axial ligands enable monooxygenase activity [47].

Surprisingly, some heme peroxidases are capable of performing not only oxidation reactions but also direct oxygen transfers (**Fig. 1.4**, *reaction 4*). This is the case of the peroxidase-peroxygenase superfamily members. These enzymes share the active site architecture of P450, endowing them with monooxygenases-like peroxygenase activity. This additional catalytic activity is not only due to the thiolate proximal iron ligand, which increases the compound II basicity due to electron push and facilitates C-H bonds oxidation [43], but also because of a more open distal active site cavity that permits a proper binding of small organic molecules and facilitates oxygen transfer [34]. The peroxygenase cycle is shown in **Fig. 1.6** (cyan-colored reaction) and exemplified with naphthalene's hydroxylation. In essence, peroxygenase mechanism is the same as the peroxidative one up to the formation of compound I. Then, the major difference is that compound I is reduced and protonated by the substrate molecule leading to compound II, that hydroxylates the substrate radical returning to the ground state [47-49].

1.2.1 Dye-decolorizing peroxidases (DyPs)

Dye-decolorizing peroxidases (DyPs, EC 1.11.1.19) are heme-containing enzymes that catalyze the one-electron hydrogen peroxide-mediated oxidation of substrates (**Fig. 1.6**, purple-colored; **Fig. 1.4B**, *reaction 1*). However, their characteristics, including amino acid sequence, distal and proximal binding site residues, tertiary structure, substrate specificity and optimum reaction conditions (pH, pressure, and temperature), differ substantially to all hitherto described peroxidases families [50, 51]. Because of that, they have been recently classified into a separate family of enzymes and grouped into the peroxidase-chlorite dismutase superfamily because of their structural resemblance with chlorite dismutases [31].

Unlike well-known peroxidases, which are primarily α-helical proteins, DyPs are structurally arranged into two domains containing α-helices and anti-parallel β-sheets, adopting a ferredoxin-like fold [52]. Furthermore, their optimal pH is quite acidic (pH 3.2) [50], and they show considerably good stability at extreme pressures and temperatures [53]. DyPs' distal binding site conserved residues are also different from those of other peroxidases. DyPs have a histidine as fifth heme ligand that is hydrogen bonded to an aspartate [54]. In *Aau*DyP, the entrance to the heme pocket is defined by the catalytic acid-base pair (Asp168 and Arg332) along with two hydrophobic residues (Leu357 and Phe359). **Fig. 1.7** shows a representation of the *Aau*DyP heme-binding pocket. The marked flexibility of the catalytic aspartate, which differs from the typical histidine shown in classic heme peroxidases, is of crucial importance for compound I formation [55] as well as for the modulation of small organic molecules entrance to the binding site [51].

Moreover, DyPs show activity towards a broad palette of substrates, including various dyes, compounds, and carotenoids [38]. But more importantly, they can oxidize substrates that are too large to fit on their active site, suggesting a LRET pathway mechanism between the heme cofactor and surface residues [54]. *Aau*DyP is particularly rich in solvent-exposed tryptophan, and tyrosine residues, and recent studies have shown its ability to oxidize bulky substrates at Tyr337 residue [56]. However, alternative LRET routes have been proposed [57] suggesting that the *Aau*DyP surface oxidation capabilities are still to be unraveled. **Fig. 1.7** shows possible surface-exposed catalytic residues in *Aau*DyP.
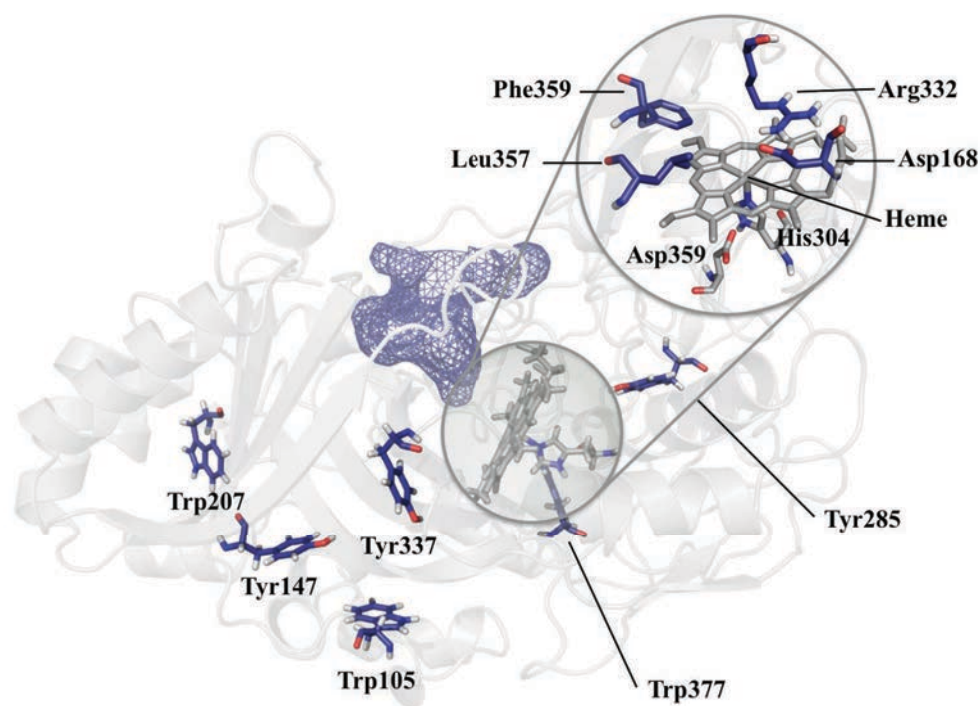
Figure 1.7. Three-dimensional structure of *Aau*DyP (PDB: 4W7J) showing solvent exposed tryptophan and tyrosine residues (Trp105, Tyr147, Trp207, Tyr285, Tyr337 and Trp377), the heme pocket entrance path (represented as a purple wireframe) and a highlight of the heme-binding site with important distal and proximal residues.

1.2.2 Unspecific peroxygenases (UPOs)

In 2004, a novel HTP type was described from the basidiomycete *Agrocybe aegerita* (*Aae*UPO) [58]. First misclassified into an alkaline lignin peroxidase, and then as a haloperoxidase and an aromatic peroxygenase (APO), it is currently classified as an unspecific peroxygenase (UPO, EC 1.11.2.1) [59] and grouped under

the peroxidase-peroxygenase superfamily along with CPO [31]. UPO is now claimed to be the missing link between CPO and P450s, sharing catalytic properties of both of them. Besides typical peroxidase reactions, such as one electron oxidations (**Fig. 1.4B**, *reaction 1*; **Fig. 1.6**, purple colored), UPO also has oxygen transfer peroxygenase activity (**Fig. 1.4B**, *reaction 4*; **Fig. 1.6**, cyan colored) including aromatic, alkylic, and aliphatic hydroxylations; aromatic and aliphatic olefin epoxidations; ether cleavage; N-dealkylations; sulfoxidations; N-oxidations; and brominations (**Fig. 1.8A**) [59-61]. Furthermore, UPO has shown to have haloperoxidase (**Fig. 1.4B**, *reaction 2*) and some catalase activity (**Fig. 1.4B**, *reaction 3*) [59, 62]. In this way, UPO represents a multifunctional novel type of oxidoreductase, combining peroxidases, monooxygenases, and catalases features, showing nearly catalytic promiscuity [61].

AaeUPO shares the active site architecture with P450s and CPO, with a cysteine (Cys36) residue coordinated at the fifth position of the heme cofactor. Unlike most well-known peroxidases, it shows a deprotonated glutamate as the acid-base catalyst (Glu196) along with an arginine residue (Arg189) (See **Fig. 1.5**, left panel)**.** The entrance to the heme pocket is highly hydrophobic, containing ten aromatic residues. Among them, five are of particular interest: Phe69, Phe121, and Phe199, that orientate the substrate; and Phe76 and Phe191, that act as gates for the substrate to enter the catalytic pocket. Besides, AaeUPO also shows a Mg^{2+} binding motif near the heme pocket [31, 59, 63].

Although one could consider P450s as the formal UPO competitor regarding applicability, UPOs show some advantages over them. Unlike P450s, which need an electron-donor partner (redox cofactors or auxiliary flavoproteins), UPOs are self-sufficient enzymes only requiring H_2O_2 to be activated (**Fig. 1.8B**). Moreover, as a result of being extracellular and soluble, they are intrinsically more stable. All these advantages, along with their remarkable promiscuity (up to 300 substrates tested positive on activity), make UPO a very appealing system for its use in industrial applications [59, 64].

Both UPO and DyP families show a significant potential for their use as biocatalysts in a variety of biotechnological applications. However, because of being recently discovered enzymes, some of their characteristics are still not clear, making it necessary to further study their catalytic mechanism along with substrate range and specificity. Furthermore, to fully take advantage of the capabilities of these enzymes, sometimes it is necessary to tune them towards desired activities and properties. In this way, molecular modelling and enzyme design open the door to the application of biocatalysts in an industrial context.
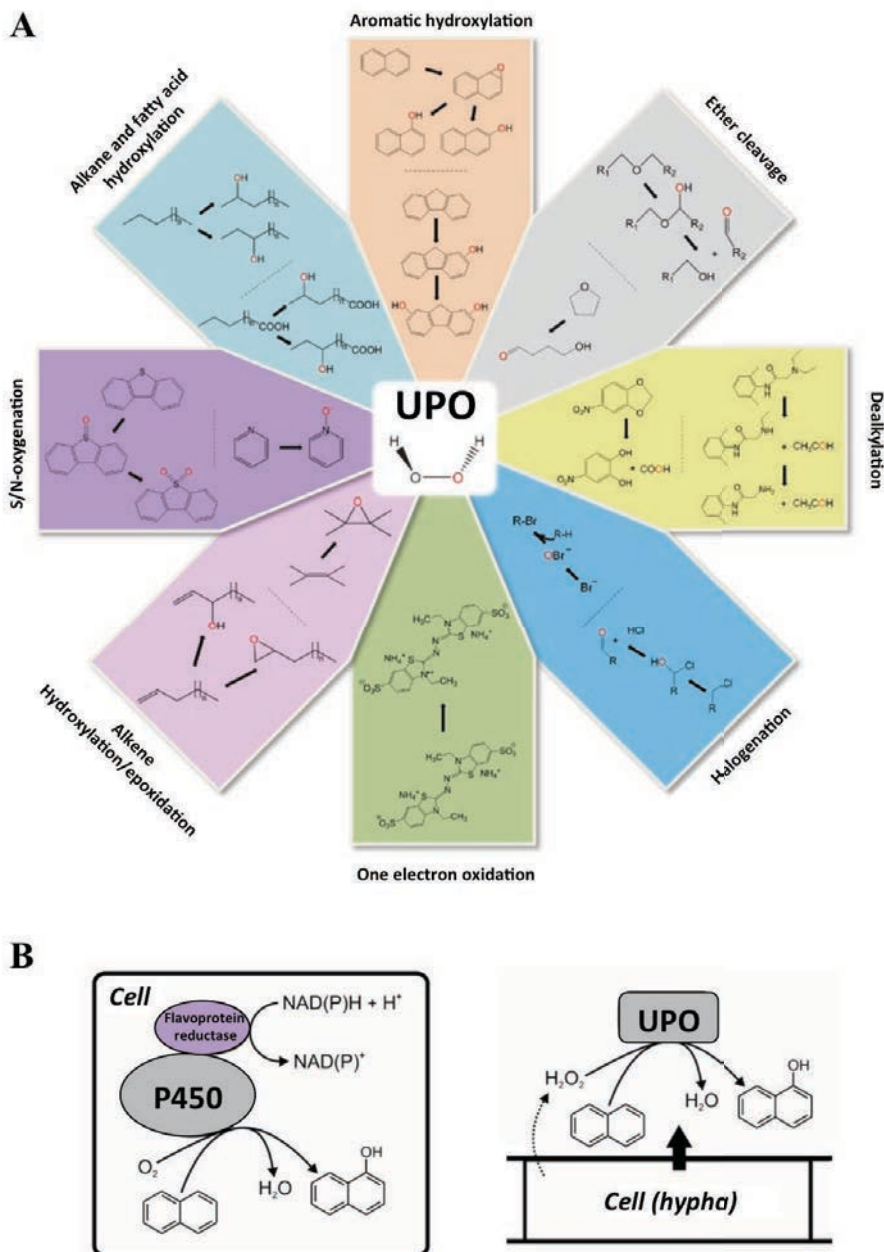
Figure 1.8. *Aae*UPO catalyzes a variety of reactions with several advantages over P450s. (**A**) Variety of UPO activities including, one-electron oxidations, halogenations and peroxygenations over different substrates. (Image retrieved from Molina-Espeja *et al.*, 2017). (**B**) Scheme representing the differences between P450s and UPO enzymes. Intracellular P450s need a source of reducing power, NAD(P)H, and an auxiliary flavin-containing reductase or protein domain while secreted UPO only requires H_2O_2 to be activated. Moreover, because of its extracellular nature, UPO enzymes are more stable [60]. (Image modified from Martínez *et al.*, 2017)

1.3 COMPUTATIONAL ENZYME MODELLING

1.3.1 Overview

Ever since they were first identified, enzymes have been the focus of many investigations in an attempt to unravel how they function, their properties and their biochemical roles. In addition to the fundamental chemical insights to be earned, such valuable knowledge can also contribute to the development of novel synthetic industrial catalysts. Experimental biochemists routinely use a variety of techniques, such as X-ray crystallography, mutagenesis studies, kinetic isotope studies and nuclear magnetic resonance, to study enzymes [65]. However, the high complexity of biocatalysts hinders in many cases the determination of their reaction mechanisms and the origin of catalysis through experiments solely [65-68]. Computational molecular modelling techniques are increasingly being used due to their unique potential to offer detailed atomic-level insights into the dynamics and reactivity of enzymes. Moreover, the exponential growth in computer power and the use of massive parallel supercomputing [69], together with the advances in theory and computer algorithms, has expanded the applicability of computational methods by improving their accuracy while enabling its application to larger systems [70]. Coupled with experimental methods, *in silico* modelling techniques can be used to unravel the mechanisms of enzyme-catalyzed reactions, identify the origins of catalysis, analyze the effects of mutations and propose new ones, and help to develop structure-activity relationships [68].

1.3.2 Methods for the *in silico* enzyme modelling

Because of the complexity underneath enzyme-catalyzed reactions, which involve a large number of processes that span a wide range of length- and time-scales, there is not a single *in silico* technique that suffices for the fully modelling of enzyme reactions. In this way, two main categories of computational methods have been developed to study enzymatic processes at a different molecular level of description: those based on classical mechanics, called molecular mechanics (MM), which seek to describe substrate migration and recognition processes, and which often involve sampling the overall protein flexibility; and those methods based on quantum mechanics (QM), which intend to explore the reaction mechanism of enzymes and that are normally focused on the binding site of the enzyme [71, 72]. Additionally, both methodologies can synergize into hybrid schemes such as combined Quantum Mechanics/Molecular Mechanics methods (QM/MM) [73].
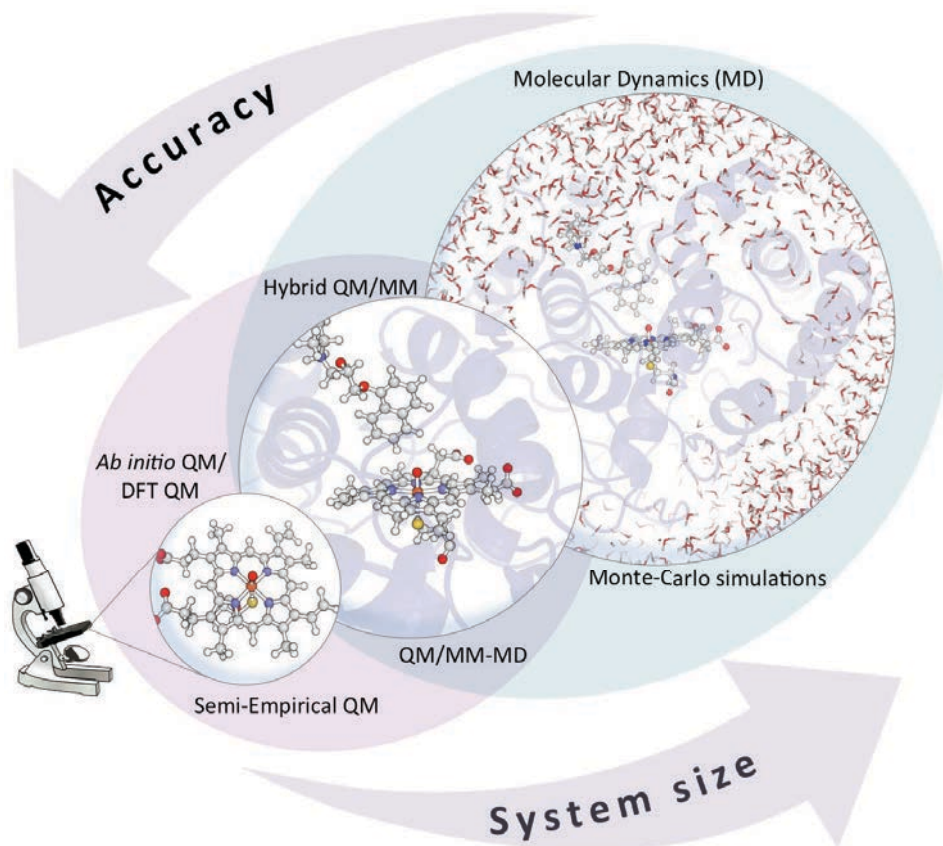
Figure 1.9. Schematic representation of some of the most used *in silico* modelling techniques. QM-based methods are circled in purple, MM-based methods are circled in blue, and hybrid methods are placed at the intersection of both. Methods are ordered upon accuracy and system size acceptance [74]. (Image adapted from Romero-Rivera *et al.*, 2017)

In this section, some of the most widely used computational methods, including those utilized in this thesis, are briefly described (see **Fig. 1.9**). Each one is different regarding the level of accuracy to describe the protein, their background theoretical algorithms and in the way of sampling the enzyme conformational space.

1.3.2.1 *Quantum Mechanics-based methods*

The modelling of chemical reactions requires a method precise enough to treat molecular systems at the detailed electronic level of theory. The most accurate theoretical framework, but also the most computationally expensive, to calculate how particles (electron and nuclei) interact within each other is QM, where the

distributions of electrons in molecules are modeled explicitly [65-68, 75]. Most QM methods ultimately trace back to the time-independent Schrödinger equation (TISE), used instead of the time-dependent equation (TDSE) when the Hamiltonian operator (Ĥ) of the system itself is not dependent on time explicitly, and the system is stationary [75-77]. TISE is the form of an eigenvalue equation that describes the quantum state of the system by means of the wave function (ψ):

$$\hat{H}\psi = (\hat{V} + \hat{T})\psi = E\psi \qquad \text{(equation 1.1)}$$

In **Eq. 1.1**, Ĥ is expressed as the sum of operators corresponding to the kinetic ($\hat{V}$) and the potential ($\hat{T}$) energies of the particles associated with the system. Ĥ acts on the wave function (ψ), with E being the energy of ψ. By solving TISE, a wave function (ψ) is obtained, which contains information about all the properties of the system [77].

However, the Schrödinger equation cannot be solved exactly for systems containing more than one electron, and thus, approximations are required to decrease TISE complexity and make it applicable to many-body systems [78, 79]. One way to simplify it is to assume the nuclei as fixed when compared to electrons. This is generally a good approximation since the nuclei, much heavier than electrons (~1800-fold), can be considered static on the electronic motion timescale. This results in the Born-Oppenheimer (BO) approximation, where the nuclei kinetic energy is not considered and leads to an "electronic" Schrödinger equation [76, 80]. Still, the electronic Schrödinger equation is intractable for many-electron systems, and further approximations are needed (such as Hartree-Fock (HF) or Linear Combination of Atomic Orbitals (LCAO) approximations) [77]. Different QM methods differ depending on the approximations applied and can be divided into three main types: *ab inito*, density functional theory (DFT) and semi-empirical approaches.

Ab initio methods

The term *ab initio* means "from the beginning" in Latin, and is given to methods that imply computations that are genuinely derived from theoretical principles, with no inclusion of empirical data [81]. In *ab initio* methods, additionally to the BO approximation, molecular orbitals are simplified by a LCAO approximation, which are ultimately defined by sets of exponents and coefficients stored in certain basis sets. The most commonly used basis sets are combinations of either Slater type orbitals (STO) or Gaussian type orbitals (GTO). The accuracy of the

calculated molecular properties will depend directly on the number of atomic orbitals and the quality of the basis set [79].

The simplest used *ab initio* QM methods apply HF approximation, in which, additionally to the BO and LCAO approximations, the Coulombic electron-electron repulsion is not taken into account explicitly, and consequently correlations between electrons are neglected [67, 76]. This approach turns out to be the main HF method shortcoming and neglecting this effect in the calculation of the total energy of the system often leads to significant errors. Both Post–HF and multi–reference methods try to solve this flaw, which results into a significant improvement in accuracy, but also to a higher computational cost. Post-HF methods correct the energy error associated with the HF approximation for a given system, the so-called electron correlation energy (E_{corr}), by means of a perturbation theory (Møller-Plesset (MP) method) or based on the variational principle (configuration interaction (CI) and coupled-cluster theory (CC) methods). In the other hand, multi-reference methods also try to correct the E_{corr} by expanding a set of Slater determinants representing excitations of the ground electronic configuration, and thus, achieving a more balanced correlation of the ground and excited states [77-79].

Density Functional Theory (DFT) methods

DFT are probably the most widely used QM methods for the study of biological systems because of their excellent compromise between accuracy and computational requirements. DFT methods rely on the Hohenberg-Kohn (HK) theorem [82], which proves that the ground-state electronic energy of a system can be defined entirely by its electron probability density (*p*) [76]. The main advantage of the HK theorem statement is that it relies on electron density properties. Since *p* is a function of only three spatial variables (*x, y, z*) that are maintained independently of the size of the system, while the complexity of the wave function (in *ab initio* methods) increases exponentially with the number of electrons. Moreover, the electron density can be measured by experimental techniques (e.g., X-ray diffraction or electron diffraction) [75]. The main problem is that the functional connecting electron density and ground-state energy is unknown. And here is where DFT methods play their part trying to connect both properties.

A practical application of this theory was developed by Kohn-Sham (KS theory), who formulated that the electron density could be expressed as a linear combination of basis functions similar in mathematical form to HF orbitals. A determinant is then formed from these functions, from which the energy of the system

can be computed [83]. The general energy functional for the KS theory can be written as in **Eq. 1.2**:

$$E[\rho(r)] = T[\rho(r)] + J[\rho(r)] + E_{XC}[\rho(r)] + E_{Ne}[\rho(r)] \qquad \text{(equation 1.2)}$$

where $E[\rho(r)]$ is the spin density energy functional; $T[\rho(r)]$ is the kinetic energy; $J[\rho(r)]$ is the Coulomb interaction energy; $E_{Ne}[\rho(r)]$ is the nuclear-electron attraction energy; and $E_{XC}[\rho(r)]$ is the so-called exchange-correlation energy [84], which is the only unknown functional.

Several functionals have been developed either from fundamental QM or by parameterizing functions to best reproduce experimental results, locating DFT in the intersection between *ab initio* and semi-empirical methods. The functional of choice will depend on several aspects, such as the target system and the computational resources available and will determine to some extent the quality of the results. As an example, in this thesis, the M06-L functional is used for the study of heme cofactor since it has been parameterized for transition metal bonding [85].

Semi-Empirical methods

Semi-empirical methods reduce the computational cost of HF calculations by decreasing the number of integrals necessary for constructing the Fock matrix. Normally, this is achieved by considering only valence electrons in the QM treatment. In this way, the matrix elements associated with the wave function interactions are not explicitly calculated via integrals but are constructed from a set of predetermined parameters. These parameters define the forms and energies of the atomic orbitals in order to reasonably agree with experimental data [76, 79, 86]. Semi-empirical methods are the less computationally intensive QM methods but are also typically the most inaccurate since the output will largely depend on the quality of the parameterization [67].

1.3.2.2 Molecular Mechanics-based methods

From Fischer's lock and key theory to explain ligand binding, our perspective on enzymatic processes has evolved considerably. Today, it is well known that protein flexibility takes an essential role in enzyme's activity, not only in the active site reorganization in ligand binding and releasing processes (Koshland's induced fit theory), but also in protein conformational fluctuations, which are known to be crucial for substrate binding, regulation, inhibition, and product release [74, 87]. However, the study of such processes, which involve large time- and length-scales,

is impractical for QM methods due to the computational costs. In this way, methods based on classical mechanics (usually referred as molecular mechanics (MM)) allow the description of larger systems (up to millions of atoms) on longer scales [88].

MM techniques rest on a view of molecules as a collection of masses interacting one with each other through harmonic forces. In this picture, balls represent atoms (all-atomistic methods) or pseudo-atoms replacing a group of atoms (coarse-grain models). Thus, the atoms or complexes of atoms are treated as balls of different sizes and flavors joined together by springs of variable strength and equilibrium distances (bonds) [89]. This simplification allows the use of MM as a fast computational model that can be applied to large biological molecules.

In MM methods, the potential energy of the system is calculated as a function of the atomic degrees of freedom, ignoring the electronic motions and implicitly assuming the BO approximation. Such functions, referred as force fields (FFs), define the strength and equilibrium values for the different possible interactions among the system and can be parameterized from experimental data or/and QM calculations [90]. Typical FFs deal with two primary classes of interactions: bonded interactions and non-bonded interactions, as shown in **Fig. 1.10**. Bonding interactions includes elements related to linked atoms such as bond, angle and dihedral terms. Stretching and bending energies are described by harmonic potentials, while torsional energies are defined by series of cosine functions. The non-bonded interactions capture longer-range interactions within the protein. The major non-bonded forces include the electrostatic interactions, based on Coulomb's law, and van der Waals interactions, based on a Lennard-Jones potential [91, 92]. In this way, the potential energy of a given system can be calculated by solving all the terms of the following equation:

$$E_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

Figure 1.10. Common formula to describe a FF, where bonding and non-bonding interactions are included. (Image retrieved from Durrant *et al.*, 2011)

Additionally, because proteins are not isolated entities, and they generally exist in an aqueous environment, it is necessary to take into consideration the treatment of the solvent. In this way, FFs can also be used to model explicitly solvent molecules. Usually, to avoid using infinite solvent molecules without having boundary effect problems, periodic boundary conditions (PBCs) are used. In PBCs, all the atoms in the system, including the solvent and the protein, are placed in a box (the unit cell) that is replicated throughout the space to form an infinite lattice where each subunit box acts as a replicate of the system in the unit cell (**Fig. 1.11**). However, the relatively high computational cost related to physically adding solvent molecules could be reduced by the use of implicit models where the solvent is treated as a structureless continuum with specific dielectric and interfacial properties, and thus, the number of degrees of freedom of the system is significantly reduced. However, the implicit atomistic description of the solvent comes with a cost: lack of solvent hydrogen bonds, and interactions within the solvent and the protein, incorrect ion distribution and unphysical sampling [93].
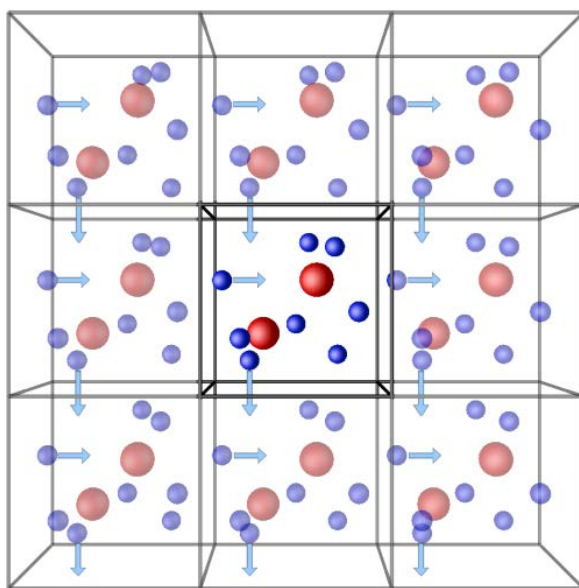
Figure 1.11. Schematic representation of the idea of PBCs. (Image retrieved from http://-isaacs.sourceforge.net/phys/pbc.html)

There are FFs for all tastes. Some of them represent all the atoms in the protein, while others treat only certain type of atoms explicitly (united-atom force-fields), others are developed specifically for particular reactions and molecules, and even polarizable FFs are being developed [94]. Nowadays, different methods to study biomolecules are based on MM and FFs, including molecular dynamics (MD), Monte Carlo (MC) and docking.

Molecular Dynamics

Among all the different MM techniques, MD is probably the most commonly used in the field of biological macromolecules. MD simulations have advanced largely since they were first developed in the late 70s, being able now to simulate systems including more than 1.000.000 atoms, membrane-embedded proteins, or large macromolecular complexes such as nucleosomes or ribosomes for long time-scales (up to the millisecond scale). This advances would not have been possible without the use of high-performance computing (HPC) and the inherent simplicity of the MD algorithm [95].
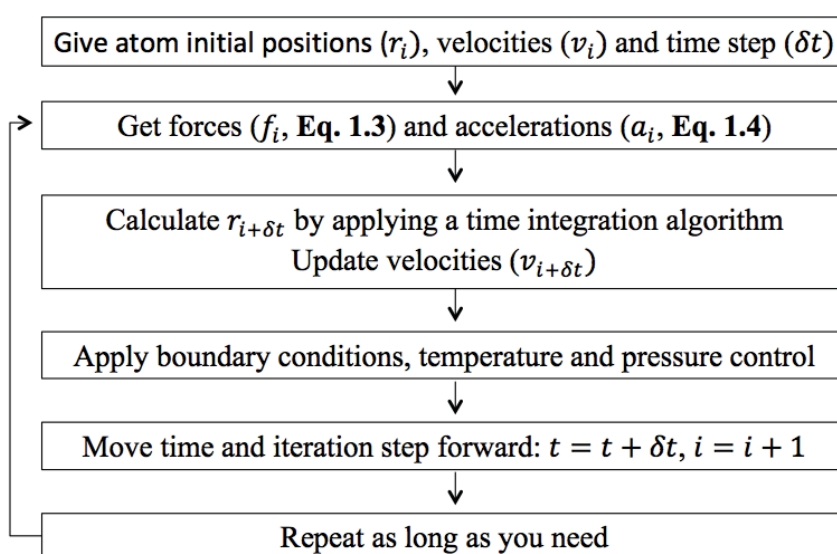
$$\boxed{\text{Give atom initial positions } (r_i)\text{, velocities } (v_i) \text{ and time step } (\delta t)}$$

$\downarrow$

$$\boxed{\text{Get forces } (f_i, \textbf{Eq. 1.3}) \text{ and accelerations } (a_i, \textbf{Eq. 1.4})}$$

$\downarrow$

$$\boxed{\begin{array}{c}\text{Calculate } r_{i+\delta t} \text{ by applying a time integration algorithm}\\ \text{Update velocities } (v_{i+\delta t})\end{array}}$$

$\downarrow$

$$\boxed{\text{Apply boundary conditions, temperature and pressure control}}$$

$\downarrow$

$$\boxed{\text{Move time and iteration step forward: } t = t + \delta t, i = i + 1}$$

$\downarrow$

$$\boxed{\text{Repeat as long as you need}}$$

Figure 1.12. Schematic representation of the algorithm followed in MD simulations.

In MD, different system configurations are obtained by the integration of Newton's second equation to propagate such system over time and let it explore the available phase space, resulting in a time-dependent trajectory. MD calculations follow a very simple algorithm, shown in **Fig 1.12**. An initial model of the system is required, along with the initial positions (r_i) and velocities (v_i) of their atoms, and the time step to be used in the simulation (δt), generally set to 2 fs. This first structure model can be obtained either from experimental structures or comparative modelling data and can be represented at different levels of granularity (all-atom or coarse-grain representation). Velocities values are usually generated according to atom types and simulation temperature. Once the system is properly solvated (typically with explicit solvent applying PBCs), forces (f_i) acting on every atom

are obtained by the use of FFs, where the potential energy $(U(r^N))$ is deduced from the molecular positions (r_i):

$$f_i = -\frac{\partial}{\partial r_i} U(r^N) \qquad (equation\ 1.3)$$

Then, classical Newton's law of motion (**Eq. 1.4**) is used to calculate accelerations (a_i) and velocities (v_i) and to update the atom positions $(r_{i+\delta t})$:

$$a_i(t) = -\frac{f_i}{m_i} \qquad (equation\ 1.4)$$

This process is repeated iteratively yielding a trajectory (**Fig. 1.12**) [96]. To achieve the efficient integration of Newton's motion law, several algorithms have been developed, being the Verlet and the Leapfrog algorithms the most widely applied.

The main bottleneck of MD is that in order to achieve the timescales of biological processes and still be able to reproduce the fastest molecular movements, very large number of calculation cycles must be done [95]. Many strategies have been developed to accelerate the simulations without simplifying the representation of the system, such as done when using coarse-grain models. These include: parallelization or the use of graphical processing Units (GPUs), the use of specific MD hardware and software (such the Anton machine [69]) and applying enhanced algorithms, such as in metadynamics, umbrella sampling or replica exchange .

Monte Carlo

As an alternative to MD simulations, in MC methods a set of non-time related representative configurations of the system are generated by a stochastic approach. The main idea behind MC is simple: random moves are added to the coordinates of the system leading to new configurations that are accepted or rejected based on a Metropolis criterion, which ensures a Boltzmann distribution of the generated ensemble [80, 96, 97]. In this way, the acceptance criterion of a step is based on the energy difference between the current and the previous configuration: if the energy decreases, the movement is accepted; otherwise, the step is accepted only when the probability of acceptance of such move (P^{acc}) is higher than a generated random number (z) on the interval [0.0, 1.0]:

$$W_{a\to b} = \begin{cases} 1, & \Delta U < 0 \\ e^{-\frac{\Delta U}{K_b T}} > z, & \Delta U > 0 \end{cases} \qquad (equation\ 1.5)$$

Here, the move acceptance probability of going from a to b molecular configuration is $W_{a \to b}$; ΔU is the potential energy difference between such configurations; K_b is the Boltzmann's constant; and T is the temperature chosen for the simulation.
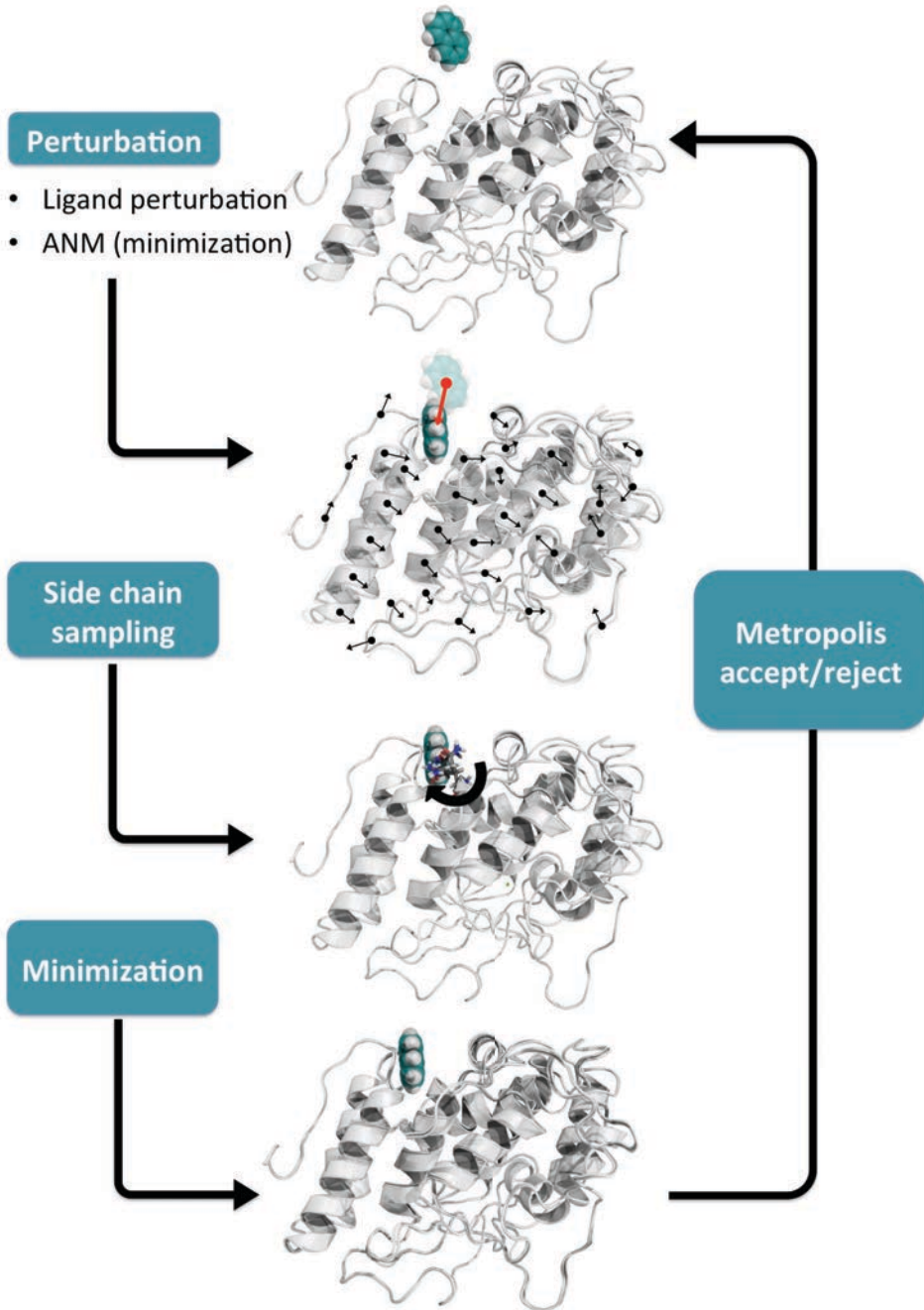

Figure 1.13. PELE algorithm scheme.

Unlike MD simulations, MC is free from the restrictions of solving Newton's equations of motion, which leads to a decrease in the computational cost while allowing the user an exceptional flexibility (non-physical moves) to approach specific problems. Moreover, the Metropolis acceptance criterion enables the possibility of accepting higher energy configurations, allowing MC methods to climb uphill and escape from local minimum. However, contrary to MD simulations, no time-related information can be gathered from traditional MC, and usually, implicit solvents are used to avoid the excessive rejection of moves due to atom overlap, which could limit the accuracy of the simulation [97].

In this thesis, MC simulations have been applied for the modelling of the protein-ligand conformational space by the use of an in-house program named PELE (Protein Energy Landscape Exploration) [98, 99]. PELE is a MC-based technique that combines protein structure prediction techniques with random system perturbation in a Metropolis acceptance scenario. PELE, initially designed for the biased mapping of ligand entrance and exit pathways [98], has been shown to accurately reproduce nonbiased ligand diffusion and active site search [100, 101], induced fit docking (IFD) [102], overall protein dynamics [103], DNA-ligand diffusions [104] and to calculate absolute binding free energies [105], at a fraction of computational resources used in MD techniques. As shown in **Fig. 1.13**, PELE's heuristic approach can be divided into two main parts: perturbation, where both protein and ligand are considered, followed by the relaxation of the system. In the perturbation step, the system is perturbed by randomly translating and rotating the ligand while the collective motion of the protein is modeled based on a combination of normal modes obtained in an anisotropic network model (ANM) [106], where the protein is represented as a network of elastic forces connecting alpha carbons. As a result, translation vectors for every normal mode are placed on each alpha carbon to describe the backbone fluctuations of the system. Moreover, the specific range of ligand's perturbation random values can be defined by the user, as well as ligand restrictions to explore particular areas of the protein (e.g., specific active site vs. overall protein surface exploration).

PELE's perturbation step is followed by the system relaxation, which includes side-chain prediction and all-atom energy minimization. Side-chain prediction involves the optimization of the side-chain rotamers (dihedral angle conformers) of those amino acid residues that are placed around the ligand or that have suffered a large energy increase as a result of the protein motion. Then, a final minimization is carried out with a truncated Newton algorithm using OPLS (Optimal Potential for Liquid Simulations)-2005 [107] FF and a generalized Born implicit solvent [108] (where explicit water molecules can also be included). The

obtained configuration is accepted or rejected based on a Metropolis criterion. The overall procedure is iterated yielding a stochastic trajectory.

Docking Methods

Although not implicitly used in this thesis, docking methods have become increasingly important tools for protein modelling, and I found necessary its inclusion in this section. Molecular docking predicts the most stable conformation of a protein-ligand complex by means of a binding free energy crude approximation. Generally, in docking approaches, random orientations of the ligand within a defined search volume of the receptor are generated and later analyzed and ranked according to a scoring function. Scoring functions can be based on empirical estimates for the interaction strengths, on MM FFs, derived from previous knowledge about possible interactions, or even derived from machine learning. Docking approaches allow the rapid evaluation of different binding modes for ligands, enabling to work with thousands of ligands, and hence, being a suitable method to be used in high-throughput virtual screening. Most widely used docking methods include Glide, Autodock, Vina, Gold or FlexX [109]. The main drawback of docking methods is the rigid treatment of the protein. Such limitation can be solved by the additional use of techniques to sample the protein backbone, such as MD or MC, to produce an ensemble of different protein conformations on which dock the ligands, or by the use of IFD procedures [110].

1.3.2.3 *Hybrid QM/MM methods*

Enzymes are too large to be treated with conventional QM methods, but at the same time, MM accuracy is not sufficient to model the electronic rearrangements occurring throughout a chemical reaction. Hybrid QM/MM methods, first introduced by Warshel and Levitt in 1976 [111], combine the strengths of both methodologies enabling the accurate study of reactive biomolecular systems at a reasonable computational cost. In this way, in QM/MM techniques, the region of the enzyme in which the chemical process takes place, commonly the active site (e.g. the reactive groups of the enzyme, substrate and any cofactors), is treated at an appropriate level of quantum chemistry theory, while the large remaining non-reactive part is described more simply by MM [112-114].

The most challenging part of the QM/MM techniques is the way the boundary region between QM and MM regions is treated. Although two different coupling strategies have been described, the subtractive and the additive coupling schemes, only the last one will be considered in this thesis. In additive schemes

(**Fig. 1.14**), the QM part is embedded within the larger MM region. Thus, the potential energy for the whole system (E) is defined as the summation of energies from both parts (E_{MM} and E_{QM}), which are described as isolated systems using the techniques discussed in previous sections, plus the energy from the frontier ($E_{QM/MM}$), which determines how these two regions interact. In this sense, several strategies have been developed to describe these interactions at various degrees of sophistication, discerning between covalent and non-covalent interactions.
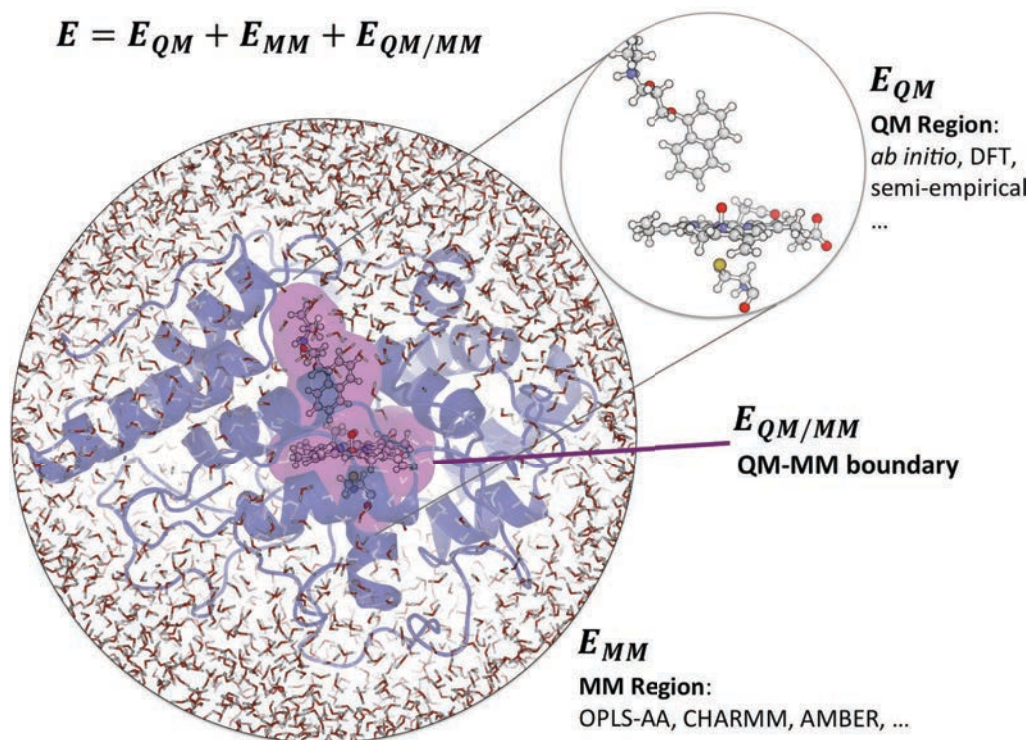

$$E = E_{QM} + E_{MM} + E_{QM/MM}$$

E_{QM}

QM Region:
ab initio, DFT,
semi-empirical
...

$E_{QM/MM}$

QM-MM boundary

E_{MM}

MM Region:
OPLS-AA, CHARMM, AMBER, ...

Figure 1.14. Schematic description of the parts involved in a QM/MM calculation [74]. (Image adapted from Romero-Rivera *et al.*, 2017)

If the interphase between the QM and the MM parts involve non-covalent interactions, different embedding approaches can be considered. In the most straightforward approach, the mechanical embedding, all the interactions between the QM and the MM systems are treated at the MM level. This treatment is appropriate for van der Waals interactions, which are hard to describe accurately with QM, but not for electrostatic ones, which often provide the dominating catalytic effect. In the electronic embedding, the electrostatic interactions between the two

systems are taken into account during the computation of the electronic wave function, enabling the polarization of the QM electronic density by the MM atom-centered point charges. Finally, higher-level corrections to include the mutual polarization between the MM and the QM regions can be included in the polarized embedding scheme [112, 115].

If there are covalent bonds between the QM and the MM systems, the description of the QM/MM frontier is more complex. In such case, several schemes have been described using link atoms or frozen orbital approaches [112]. The easier solution is to introduce a monovalent link atom (normally a hydrogen atom) at an appropriate position along the bond vector between the QM and MM atoms. A popular alternative to the link atom scheme uses frozen localized orbitals at the QM/MM boundary that can be placed on the last QM atom, as in the localized self-consistent field method, or on the MM atom, as the generalized hybrid orbital approach (GHO) [115]. In this thesis, all QM/MM calculations have been performed using the QSite software [116], which treats the QM/MM boundary with the hydrogen cap and electrostatic embedding approaches while using DFT to describe the QM system and OPLS-2005 FF in the MM region.

1.3.3 Protein-related theories

In this section, different theories that are closely-related with enzyme function and that are fundamental to understand the work presented in this thesis are introduced: enzyme kinetics, electron transfer (ET) and LRET pathways in proteins. Additionally, a general perspective relating such theories to its molecular modelling is given.

1.3.3.1   *Enzyme kinetics*

Assuming the following mechanism for a simple single-substrate enzyme chemical reaction:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \overset{k_2}{\rightarrow} E + P \quad (equation\ 1.6)$$
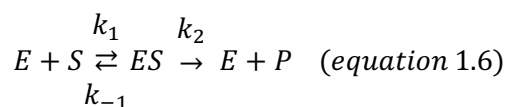
where E is the enzyme binding to the substrate, S, to form the Michaelis complex (ES), with a rate of formation k_1. Then, ES has two possibilities: to dissociate to E and S again, with a rate of dissociation of k_{-1}, or to proceed to form the product

(P) with a rate of formation of k_2. The catalytic rate of such reaction (v) can be modeled by the Michaelis-Menten equation:

$$v = \frac{v_{max}[S]}{K_M + [S]} = \frac{k_{cat}[E][S]}{K_M + [S]} \quad (equation\ 1.7)$$

Here, v_{max} is the maximum conversion rate of substrate S to product P, and is directly related with the TON of the enzyme, also called k_{cat}, which is the number of S converted into P in a unit of time when E is fully saturated with S. When v_{max} is reached, so that $[S] \gg K_M$, k_2 coincides with k_{cat}. $[S]$ is the concentration of unbound substrate, $[E]$ is the total concentration of free enzyme and K_M is the Michaelis-Menten constant, which is defined as the substrate concentration at which the reaction rate is half of v_{max} for an enzyme obeying Michaelis-Menten kinetics [117].
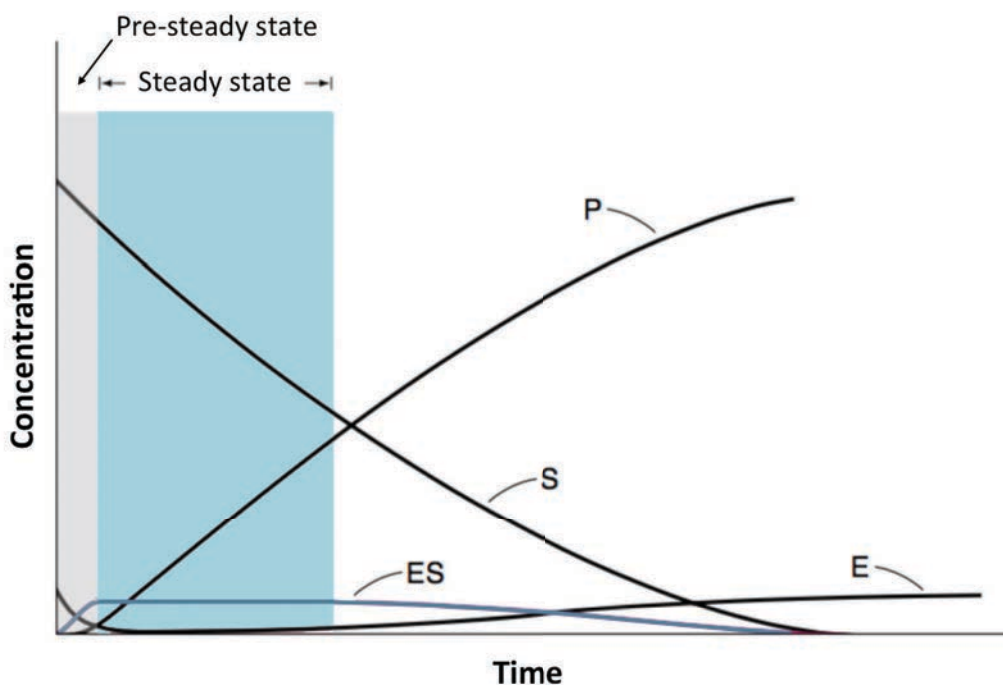

Figure 1.15. Plot of the concentrations of the various species in a Michaelis-Menten model enzymatic reaction. Pre-steady and steady states are highlighted in gray and turquoise respectively. (Image modified from http://www.uscibooks.com/changten.pdf)

In the pre-steady-state approximation, when $k_{-1} \gg k_2$ so that ES dissociates to E and S much more rapidly than P is formed (**Fig. 1.15**, gray-colored area), K_M is equal to the dissociation constant of the ES complex. When this condition is

met, K_M by itself is a measure of the strength of the ES complex: a high K_M indicates weak binding while a low K_M indicates strong binding. On the other hand, in the steady-state approximation, when $k_2 \gg k_{-1}$, the formation of P is much more faster than ES formation (**Fig. 1.15**, turquoise-colored area), and thus, $d[ES]/dt = 0$. In these conditions, K_M explicitly depends on k_{cat}:

$$K_M = \frac{k_{cat} + K_{-1}}{K_1} \quad (equation \ 1.8)$$

Such approximations would facilitate to a considerable extent the computation of kinetic constants. However, we have to take in mind that none of these assumptions would be valid in a real enzyme reaction framework, and that, besides, the rate of product dissociation should also be considered in those circumstances where the product disassociation is slow. In this way, the K_M estimation by molecular simulations is very challenging as it potentially depends on the three rate constants in **Eq. 1.8** (in addition to product dissociation if needed). Only if P formation is much slower than ES formation, K_M is equal to the dissociation constant of ES, and its computation would be directly through protein-ligand binding modelling. On the other hand, k_{cat} can be reduced to the calculation of the rate constant according to Eyring's transition state (TS) theory or by computing key parameters of Marcus' equation (**Eq. 1.9**) if ET occurs.

1.3.3.2   *Electron transfer in proteins*

The ET process is one of the most ubiquitous and fundamental phenomena in a diverse array of biological transformations, ranging from photosynthesis to aerobic respiration [118]. Its appearance in heme peroxidases processes is not an exception, where ET is used to oxidize a wide variety of substrates [34]. ET processes might involve a relatively short pathway, e.g. from a cofactor or substrate directly bound in the vicinity of the acceptor group, or LRET pathways, where the donor and the acceptor might be substantially apart from each other, e.g. across protein-protein complexes or between the cofactor and surface residues [119, 120], as seen in some heme peroxidases [51].

Electron transfer mechanism

The ET process in proteins can be described by the Marcus theory [121] where an electron from a donor (D) species is transferred to an acceptor (A), so the reactant (R) state is D^-A, and the product (P) state is DA^-. When there is a weak coupling between D and A species, as in biological systems, the ET is a non-

adiabatic process described by the rate constant (k_{ET}), which can be derived from the Marcus equation:

$$k_{ET} = \frac{2\pi}{\hbar} |V_{DA}|^2 \frac{1}{\sqrt{4\pi\lambda k_B T}} e^{\left[-\frac{(\Delta G° + \lambda)^2}{4\lambda k_B T}\right]} \quad (equation\ 1.9)$$

where $\hbar$ is the reduced Planck's constant, k_B the Boltzmann's constant and T the temperature. According to Marcus theory, the three resting parameters determine the rate of ET in proteins: (1) the overall Gibbs free energy change of the ET reaction ($\Delta G°$); (2) the energy required to move the nuclear coordinates from R to P state prior to ET (λ, also named reorganization energy); and (3), the electronic coupling between D and A (V_{DA}). $\Delta G°$ is the difference in free energy between R and P states, and thus, it affects the equilibrium of the reaction. The other two properties influence the rate of the reaction. In particular, V_{DA} is the most geometrically dependent element in the kinetic constant since its value depends on the coupling matrix when D and A are in contact, and in the distance between them (V_{DA} decays as distance increases) [118].

The rigorous theoretical solving of the ET rate (k_{ET}) is far from trivial since it involves all **Eq. 1.9** parameters, and, although $\Delta G°$, λ and V_{DA} are computationally accessible, its calculation requires a significant effort. As an alternative, the electron coupling (V_{DA}), which is quadratically proportional to the ET rate, is often used as an easier and faster way to estimate ET propensities [122].

Long-range electron transfer (LRET) pathways

As already mentioned, ET can also occur even if the donor and the acceptor species are set far away by a LRET mechanism. The long-range electron tunneling can be treated in a weak coupling or non-adiabatic transfer, due to large distances between the involved species, and Marcus formula (**Eq. 1.9**) can be applied [123]. In such situation, the detailed structure of the protein medium intervening between the redox centers, or in other words, the residues involved in the ET, become very important regarding a better understanding of the enzymatic cycle. Additionally, such information is very valuable to guide enzyme engineering studies [120].

The identification of key intermediate residues along the ET pathway is usually accomplished by means of site-directed mutagenesis. However, apart from being a time- and money-consuming technique, mutations might cause structural changes or significant protein electrostatic perturbations, making difficult to reveal the real cause for the electron conductivity loss. In this respect, computational tools

offer an alternative approach to describe at the atomic and electronic level electron pathways. Currently, several programs based on topological search algorithms to find the shortest pathway or on semi-empirical extended Hückel level of theory, are able to obtain an approximated ET pathway [120].

In this thesis, an *ab initio* based computational program, named QM/MM e-Pathway [120], has been used to map LRET pathways. This method makes use of the different levels of theory employed in QM/MM to rank specified residues upon their tendency to host the radical formed during the ET process. The method's strategy is based on modifying the residues included in the QM and MM region to follow the evolution of the spin density and underline the pathway taken by an electron from the donor to the acceptor. QM/MM e-Pathway procedure scheme is shown in **Fig. 1.16**.
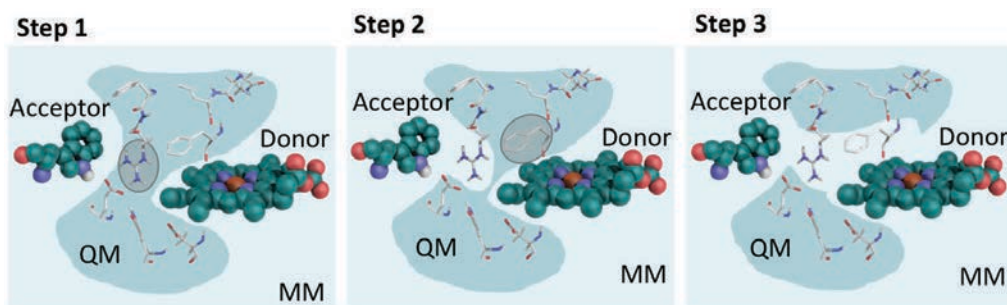

Figure 1.16. Schematic representation of the QM/MM e-Pathway method. The protein transfer region is treated at the QM level of theory (colored in dark turquoise) while the rest of the protein is treated with MM (colored in light turquoise). Once an amino acid is identified as a spin residue host (highlighted in gray) it is removed from the QM region (step 2), and spin density is computed again. The search is iterative until the path that links the donor and the acceptor is found (step 3).

1.4   COMPUTATIONAL ENZYME ENGINEERING

1.4.1 Overview

As already highlighted, most enzymes are tremendously proficient catalysts that can be used in an extracellular medium for a whole host of processes. However, naturally occurring enzyme's properties are usually not optimal for the actual industrial and biotechnological needs, and methods able to shape their physico-chemical properties effectively are usually needed. Moreover, it would be ideal to be able to drive them towards catalyzing totally new reactions or to directly create from scratch new catalysts holding the desired properties [25].

Directed evolution (DE), where the enzyme is iteratively mutated aiming to improve a property of interest, has proven to be a powerful tool to "evolve" the enzyme towards wider applications. However, despite the enormous power of this approach and its relation to the advances in protein engineering, it still has several limitations, especially related to the sequence space that can be screened. In this regard, computers have the potential to play a major role in guiding and directing enzyme engineering's experimental work. The molecular modelling methods explained in the previous section can be used synergistically with experimental data in order to reduce the number of variants to be characterized experimentally, and thus, saving resources and time compared with wet-lab only studies.

Although the work performed in this thesis is more centered in enzyme molecular descripting, I found essential to introduce the methodology used for the *in silico* engineering of enzymes. In this sense, enzymes can be completely designed from scratch (*de novo* design) or modified using as starting point already existing enzymes so as to impart novel properties on them (enzyme redesign).

1.4.2 Computational *de novo* enzyme design

Computational *de novo* design involves creating a new enzyme with the desired activity from scratch, by introducing amino acid residues essential for catalysis into existing scaffolds. The underlying idea, based on Pauling's principle, is that if the active site has an optimal structural organization to sufficiently stabilize the TS of a certain reaction, such reaction will potentially occur [124]. Two of the most widely known programs to perform *de novo* design are Rosetta software suite [125], developed in David Baker's laboratory, and ORBIT [126] from Stephen Mayo.

All *de novo* design programs involve the same general scheme (**Fig. 1.17**). Initially, the TS of the desired reaction and the idealized active site geometry are modeled using QM. This composite, named "theozyme", is then docked into different protein scaffolds, and resulting complexes are optimized *in silico* by a combinatorial search for amino acid substitutions that improve TS stabilization. Finally, the designs are filtered on the basis of TS binding energy and catalytic geometry, and the top-ranked designs are synthesized and characterized experimentally [124, 127]. Moreover, computational techniques can be used to design not only the enzyme's active site but also new functional protein scaffolds from the ground up [128, 129].

Although *de novo* design comprises a very promising approach, designed enzymes still show poor rate enhancements when compared to naturally occurring biocatalysts. Such discrepancy is a result of the contribution of different structure and mechanistic factors: a suboptimal binding pocket's shape and electrostatic complementarity to the TS; to consider only the TS binding important for catalysis without taking into account other crucial states such as substrate binding and substrate/product unbinding; the non-consideration of the in/out substrate and product paths; and disregarding the scaffold flexibility [130]. In this respect, distinct strategies are commonly used to further improve *de novo* designed enzymes such as using molecular modelling to guide their redesign or DE to achieve the desired properties.
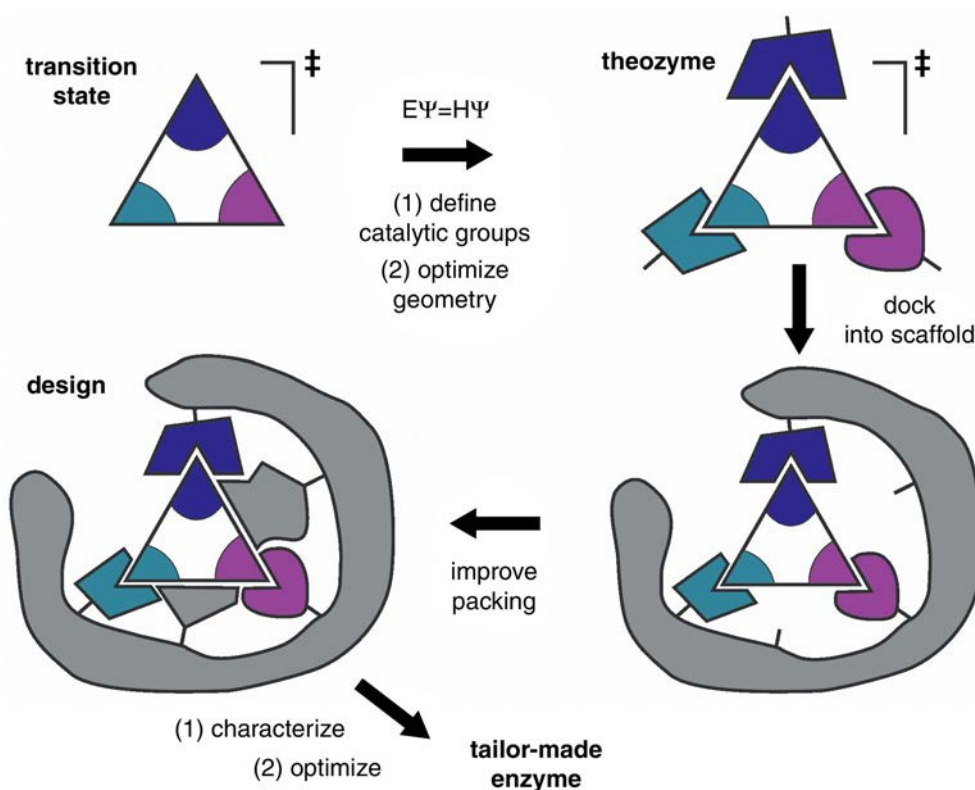

Figure 1.17. General *de novo* enzyme design schema. (Image retrieved from Kries *et al.,* 2013)

1.4.3 Computational enzyme redesign

Although less ambitious than full *de novo* design, the redesign of already existing enzymes is the most direct route for the tailoring of biocatalysts with industrial applications. In an *in silico* enzyme redesign, a structural model of the

natural enzyme, obtained ideally from a high-resolution PDB (Protein Data Bank) crystal structure or, if not available, from a homology model, is mutated in a process that can involve various degrees of rationality.

1.4.3.1 *Rational redesign*

Rational redesign usually requires not only the availability of a structural enzyme model but also knowledge about the relationships between sequence, structure and mechanism/function, and is therefore very information-intensive. On the one hand, the increasing availability of structural data in public protein databases has eased to obtain an enzyme structural model. In the other hand, with the help of *in silico* molecular modelling techniques valuable mechanistic and structural information can be obtained and used to propose mutations to increase ligand binding and the catalytic rate constant [25, 131]. Although many computational approaches exist to guide rational redesign, such as sequence-based strategies or methods that use statistical data, these are out of the scope of this thesis, and only physics-based methods will be described.

Ligand binding redesign

The protein-ligand recognition process is crucial for enzyme's activity since is the first step in enzyme catalysis, and its modelization holds a huge potential to evaluate and design enhanced mutants. However, the design of such process faces several challenges including sampling of the enormous possible orientations of the ligand with respect to the protein, the large conformational and sequence space of the protein active site (for 10 residue positions there exist ~10^{13} sequences), and the difficulties involving binding energies calculation [132]. Despite the challenges, several approaches have been developed to model and design protein-ligand binding.

Among all the computational methods to redesign protein-ligand binding, Rosetta [133] and OSPREY (Open Source Protein REdesign for You) [134] suite of programs are the most widely used. The Rosetta protocol involves the optimization of the ligand placement in the binding pocket, followed by optimization of side-chain identity and conformation [135]. OSPREY identifies protein mutants that possess desired target properties. In addition, it incorporates a set of several different algorithmic modules for structure-based protein design, including a number of powerful Dead-End Elimination (DEE) algorithms and the ensemble-based *K** algorithm for improved protein-ligand binding prediction.

On the other side, a wide variety of computational methods are used not to directly redesign, but to evaluate, filter, and correct designs. Based on FFs, docking is a cheap and fast methodology to evaluate significant interactions between the protein and the bound ligand. Thereby, docking methods (including IFD) are widely used to evaluate and optimize enzyme designs. However, even in the protein-ligand binding process, enzymes cannot be considered as static entities. Thus, taking into account the flexibility not only of the active site side-chains but also of the overall protein' backbones is essential to achieve a certain degree of accuracy. In this regard, methods based on MD, such as high-throughput molecular dynamics (HTMD) and steered MD, can be used to filter and correct enzyme designs [136, 137]. With such techniques, many features can be investigated, including binding site pre-organization, hydrogen bond geometries, binding site structural integrity, solvent exposure [26], binding site pre-organization effects caused by distant mutations [138] and the ligand entrance/exit path. As an alternative to MD, MC-based techniques like PELE software [98] can also be used to examine and rank enzyme designs with reasonable good accuracy and considering protein flexibility. Moreover, MC-based methods are less computationally expensive allowing an extensive sampling of the sequence space.

Enzymes have complex ligand migration pathways and hidden active sites, and thus, localizing tunnels and cavities that might play a role in enzyme function is important. In this way, additional tools, which do not explicitly model neither ligand sampling nor protein dynamics, can be used to identify and evaluate enzyme's tunnels and cavities. Such tools include: CAVER [139] and MOLE [140], for the identification of tunnels among the enzyme; and Fpocket [141] and PASS [142] for the computation of cavities volumes.

Catalytic rate constant redesign

The catalytic rate constant is claimed to describe the rate-limiting step of a certain reaction, and thus, its modelling and design are of particular importance for the tailoring of enzymes. Its improvement involves the modelling of the TS of the slowest chemical step of the reaction, which is far from trivial.

Several procedures exist for the redesign of enzymes using transition state analogues (TSA) or models, such as OptZyme [143] or using RosettaDesign to stabilize a TS model [144]. These methods are based on optimizing the classical interaction energy between the protein and the bound TS model upon sequence search, and although they are very useful when a big number of mutants have to be screened, they neglect the TS electronic description.

The proper computation of the activation energy barriers is a complex matter and requires an electronic description of the TS. In this way, QM/MM methodologies can be employed to redesign enzymes according to their protein-TS interaction energies [145]. However, these techniques are computationally demanding, and its application for the screening of a large number of variants is unfeasible. As a cheaper alternative to QM/MM methods, empirical valence bond (EVB) or semi-empirical methods can be applied to score a more significant number of mutants at expenses of accuracy.

In the particular case of proteins that involve ET processes, such as oxidoreductases, additional complexity is added due to the dependence of the ET rate on the electronic coupling, the reorganization energy and the driving force (see **Eq. 1.9**). In an attempt to reduce the computational cost required to study ET rates and make it applicable to the enzyme redesign problem, ET evaluation can be simplified by means of spin transference [146] or electronic coupling calculations [71]. Moreover, key mutations along LRET pathways may be identified and evaluated by the use of methodologies such as QM/MM e-Pathway [119].

1.4.3.2   *Semi-rational redesign*

Semi-rational redesign approaches utilize information on protein sequence, structure, and function, as well as computational predictive algorithms to limit the protein region and the amino acid diversity to be explored in enzyme engineering studies [147, 148]. In this way, rational and random tools are combined to create small "smart" libraries, which dramatically reduce the sequence space to be investigated and increase the success rate of finding favorable mutations. Semi-rational approaches are especially advantageous when no high-throughput assay system, such as DE, are available.

In this thesis, a computational protocol developed in our group has been used for the massive residue mutation towards an enhanced activity, in an *in silico* DE fashion [26]. Ideally, this approach should be synergically combined with experimental methods in an iterative workflow where, at each round, mutations are added. The *in silico* enzyme evolution protocol involves, first, a scrupulous selection of a ligand initial structure, which must embody a catalytic substrate positioning, and that will determine the outcome of the computational design. Then, all those residues located close to the bound substrate are identified and mutated to all or to a subset of amino acid residues. Finally, such mutants are screened with PELE software using the protein-substrate interaction energy after and induced fit procedure. Best mutants are identified, and ideally, tested with an experimental

procedure. **Fig. 1.18** shows a general schema of the *in silico* DE protocol. This approach allows quickly guiding experimental mutagenesis since a high number of enzyme positions can be scored in a short time (with 100 CPUs (central processing units), ~200 positions/day) [26].
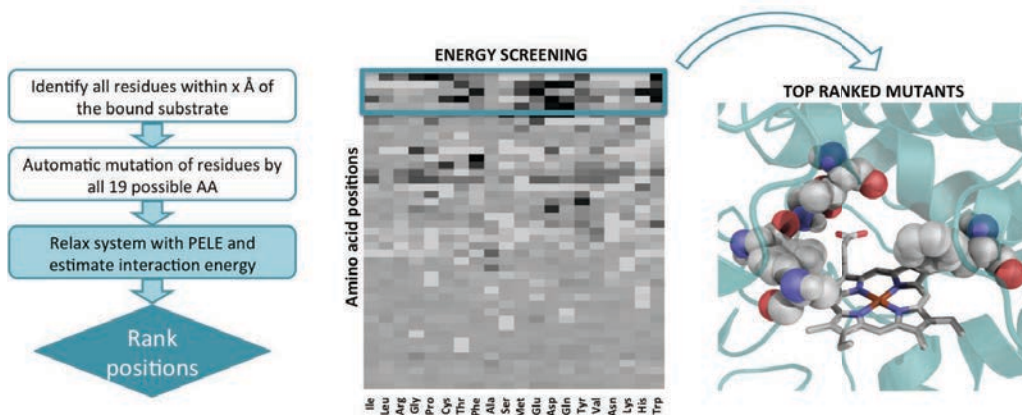

Figure 1.18. Left panel - General scheme of the *in silico* enzyme evolution protocol. **Center panel -** Heat-map of tested mutations in different residue positions. The darker the color, the higher is the mutant scoring when compared to the wild-type (based on protein-substrate interaction energy). From the energy screening, best mutants or residues positions can be identified (**right panel**) and they are further suggested to be tested experimentally. (Image modified from Monza *et al.,* 2017)

Objectives

2

"Lots of things are mysteries. But that doesn't mean there isn't an answer to them. It's just that scientists haven't found the answer yet.

For example, some people believe in the ghosts of people who have come back from the dead. […] Eventually scientists will discover something that explains ghosts, just like they discovered electricity, which explained lightning, and it might be something about people's brains, or something about the earth's magnetic field, or it might be some new force altogether. And then ghosts won't be mysteries. They will be like electricity and rainbows and non-stick frying pans.

But sometimes a mystery isn't a mystery. And this is an example of a mystery which isn't a mystery."

Mark Haddon, *The curious incident of the dog in the night-time*

As highlighted in the previous section, *in silico* methodologies have progressively turned into highly valuable tools for the study and design of enzymatic systems due to their unique potential to offer atomic-level insights into biocatalysts' dynamics and reactivity. Furthermore, the exponential growth in computational power and parallelized calculations on graphical cards, along with the continuous algorithms improvements, are continuously expanding their applicability to larger systems and improving their accuracy. Over the last few years, computational approaches have been successfully complemented in an interactive manner with experimental techniques, such as DE or site-directed mutagenesis. In this way, cleverly combined computational methods can provide information about enzyme-catalyzed reactions that experiments cannot, shedding light on catalysis specificity and efficiency determinants and opening the door to the design of new synthetic biocatalysts with potential value for industrial and academic applications [68, 149, 150].

Motivated by the advances on computational techniques and by the ease of obtaining valuable experimental data, which has been provided by our collaborators, the main goal of this thesis is to understand, at the atomic level, the mechanisms of reaction of the peroxidases under study here by using *in silico* methodologies. Moreover, by utilizing the acquired knowledge, we aim to evaluate experimentally obtained enzyme variants and to guide the design of new ones towards desired properties. Since this thesis is presented as a compendium of publications, I have classified each work according to the question that aims to be answered rather than the enzyme system that is studied. In this way, each section of **Chapter 3 - Results**, will include a number of publications (either published or in process to be published) attempting to achieve the below described objectives: i) Reaction mechanisms description, ii) Enhanced variant rationalization, and iii) Computational enzyme redesign. Besides, the specific goals of the here presented thesis are the following:

i. **Reaction mechanisms description.** The first specific objective seeks to unravel the mechanistic details regarding alternative oxidation sites in peroxidases, in particular in *Aau*DyP, which has been suggested to involve a LRET route from radical-forming aromatic residues at the protein surface [56]. On the other hand, studying the mechanistic aspects of hydroxylation and oxidation processes in peroxidases is crucial both for a better understanding of enzyme's activity and for the design of enhanced biocatalysts. In this way, we aim to shed light on the determinants behind catalytic efficiency (involving K_M and k_{cat}) and to provide insights into the factors that guide regioselectivity in UPO enzymes.

ii. **Enhanced variant rationalization.** By the application of several *in silico* approaches, which provide information at different levels of accuracy, along with the use of previously acquired knowledge about enzyme reaction mechanisms, we aim to rationalize the effect of mutations on DyP and UPO enzymes' activity by evaluating experimentally obtained variants.

iii. **Computational enzyme redesign.** Based on the experience and knowledge accumulated on the way to achieve the previously mentioned objectives, we intend to make a step forward and move from the molecular description of biochemical and biophysical processes to the actual prediction of such processes. In other words, we aim to *in silico* engineer enzymes towards desired properties for their potential use in industrial applications. In particular, a new computational semi-rational approach will be applied to engineer *Aae*UPO for the efficient biocatalytic synthesis of high-value chemicals.

Results

3

*"TRY
EXPERIMENT
ITERATE
FAIL
TRY AGAIN"*

Simon Sinek

3.1   REACTION MECHANISMS DESCRIPTION

A variety of processes, spanning in wide-range in time- and length-scale, are involved in enzyme-catalyzed reactions. For an isolated system, these processes include: i) substrate binding and substrate/product unbinding to the enzyme's active site, ii) in/out substrate and product migration paths, iii) conformational exchange between various sub-states during the catalytic cycle and iv) chemical catalytic steps involving the breaking and forming of bonds and/or the transfer of electrons [74, 151]. In this way, the use of molecular modelling techniques can be of special importance when trying to understand biomolecules. Because of this diverse number of processes involved in enzyme catalysis, there is not a single theoretical method that suffices for a complete reaction description, and a variety of approaches, going from MM to QM, have to be utilized. In the present thesis section, the use of various *in silico* techniques has been crucial to unravel the atomic and electronic details under peroxidases mechanisms.

DyPs are capable of oxidizing substrates that are too bulky to be accommodated in their heme-binding site, suggesting the existence of LRET routes between the heme cofactor and surface aromatic residues [56]. In *Article I*, LRET routes are computationally studied in *Aau*DyP enzyme. In this way, PELE simulations, used to describe the substrate binding at surface interaction sites, along with subsequent QM/MM analyses, which indicated the preference of substrate oxidation at those interaction sites, are combined to obtain an atomic and electronic description of the process. Computational results, together with experimental data provided by our collaborators, allowed us to clarify the LRET mechanisms and to explain the role of protein radicals in DyP catalysis.

Computational approaches can be used to explain enzyme-catalyzed reactions in a way that experiments cannot, providing atomic/electronic insights into the molecular factors that govern yield and selectivity in biocatalysts. In *Article II* and *Article III,* the PELE software is used to simulate the diffusion and binding of high-value compounds (including steroids and vitamin D derivatives) to UPO enzymes (*Aae*UPO and *Cci*UPO). Appropriate analyses of the resulting simulations can provide rational atomic-based explanations of the experimentally obtained hydroxylation conversion rates and regioselectivities.

Publications presented in this section:

Article I: <u>Cañellas M.</u>, Linde D., Pogni R., Lucas F., Guallar V., Barratto M.C., Sinicropi A., Sáez-Jiménez V., Coscolín C., Romero A., Medrano F.J., Ruiz-Dueñas F.J. and Martínez A.T. (2015) Catalytic surface radical in dye-decolorizing peroxidase: a computational, spectroscopic and site-directed mutagenesis study. *Biochemical Journal.* 466: 253-262.

Article II: Babot E.D., del Río J.C., <u>Cañellas M.</u>, Sancho F., Lucas F., Guallar V., Kalum L., Lund H., Gröbe G., Scheibner K., Ullrich R., Hofrichter M., Martínez A.T. and Gutiérrez A. (2015) Steroid Hydroxylation by Basidiomycete Peroxygenase: a Combined Experimental and Computational Study. *Applied and Environmental Microbiology.* 81(12): 4130-4142.

Article III: <u>Cañellas M.,</u> Lucas F., Babot D.E., del Río J.C., Kalum L., Ullrich R., Hofrichter M., Guallar V., Martínez A.T. and Gutiérrez A. (2016) Molecular determinants for selective C_{25}-hydroxylation of vitamins D_2 and D_3 by fungal peroxygenases. *Catalysis Science & Technology*. 6:288-295.

3.1.1 *Article I:* Catalytic surface radical in dye-decolorizing peroxidase: a computational, spectroscopic and site-directed mutagenesis study.

Dolores Linde*[1], Rebecca Pogni†[1], <u>Marina Cañellas</u>‡§[1], Fátima Lucas‡, Victor Guallar‡||, Maria Camilla Baratto†, Adalgisa Sinicropi†, Verónica Sáez-Jiménez*, Cristina Coscolín*, Antonio Romero*, Francisco Javier Medrano*, Francisco J. Ruiz-Dueñas*[2] and Angel T. Martínez*[2].

*Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain

†Department of Biotechnology, Chemistry and Pharmacy, University of Siena, I-53100, Siena, Italy

‡Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain

§Anaxomics Biotech, Balmes 89, E-08008 Barcelona, Spain

||ICREA, Passeig Lluís Companys 23, E-08010, Spain

[1]These authors contributed equally to this work.

[2]To whom correspondence should be addressed (email **fruiz@cib.csic.es** and ATMartinez@cib.csic.es).

Catalytic surface radical in dye-decolorizing peroxidase: a computational, spectroscopic and site-directed mutagenesis study

Dolores Linde*, Rebecca Pogni†[1], Marina Cañellas‡§[1], Fátima Lucas‡, Victor Guallar‡||, Maria Camilla Baratto†, Adalgisa Sinicropi†, Verónica Sáez-Jiménez*, Cristina Coscolín*, Antonio Romero*, Francisco Javier Medrano*, Francisco J. Ruiz-Dueñas*[2] and Angel T. Martínez*[2]

*Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain
†Department of Biotechnology, Chemistry and Pharmacy, University of Siena, I-53100, Siena, Italy
‡Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain
§Anaxomics Biotech, Balmes 89, E-08008 Barcelona, Spain
||ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain

Dye-decolorizing peroxidase (DyP) of *Auricularia auricula-judae* has been expressed in *Escherichia coli* as a representative of a new DyP family, and subjected to mutagenic, spectroscopic, crystallographic and computational studies. The crystal structure of DyP shows a buried haem cofactor, and surface tryptophan and tyrosine residues potentially involved in long-range electron transfer from bulky dyes. Simulations using PELE (Protein Energy Landscape Exploration) software provided several binding-energy optima for the anthraquinone-type RB19 (Reactive Blue 19) near the above aromatic residues and the haem access-channel. Subsequent QM/MM (quantum mechanics/molecular mechanics) calculations showed a higher tendency of Trp-377 than other exposed haem-neighbouring residues to harbour a catalytic protein radical, and identified the electron-transfer pathway. The existence of such a radical in H_2O_2-activated DyP was shown by low-temperature EPR, being identified as a mixed tryptophanyl/tyrosyl radical in multifrequency experiments. The signal was dominated by the Trp-377 neutral radical contribution, which disappeared in the W377S variant, and included a tyrosyl contribution assigned to Tyr-337 after analysing the W377S spectra. Kinetics of substrate oxidation by DyP suggests the existence of high- and low-turnover sites. The high-turnover site for oxidation of RB19 ($k_{cat} > 200$ s^{-1}) and other DyP substrates was assigned to Trp-377 since it was absent from the W377S variant. The low-turnover site/s (RB19 $k_{cat} \sim 20$ s^{-1}) could correspond to the haem access-channel, since activity was decreased when the haem channel was occluded by the G169L mutation. If a tyrosine residue is also involved, it will be different from Tyr-337 since all activities are largely unaffected in the Y337S variant.

Key words: catalytic protein radical, dye-decolorizing peroxidase, EPR spectroscopy, molecular docking, QM/MM, site-directed mutagenesis.

INTRODUCTION

DyPs (dye-decolorizing peroxidases) (EC 1.11.1.19) represent a new family of haem peroxidases widespread in bacteria, archaea, fungi and other micro-organisms [1–4]. Among those of fungal origin, the enzymes from *Bjerkandera adusta* [5–7] and *Auricularia auricula-judae* [8–10] have been crystallized and biochemically characterized as representative DyPs from two phylogenetically different basidiomycetes (in orders Polyporales and Agaricales respectively). The structures of bacterial DyPs were simultaneously solved [11–15]. *B. adusta* DyP was largely characterized as a recombinant protein [16,17], whereas *A. auricula-judae* DyP was isolated from fungal cultures [18]. The latter enzyme has recently been overexpressed in *Escherichia coli* as inclusion bodies, and a refolding protocol was optimized yielding a recombinant DyP with basically the same properties as those of wild-type DyP [19].

Xenobiotic anthraquinone-type dyes are the best-known substrates for DyPs. Among wood-rotting basidiomycetes, DyP genes are significantly more frequent in the sequenced genomes of white-rot (ligninolytic) than brown-rot species [20]. This fact and their reported capability to degrade non-phenolic lignin model dimers, although with much lower efficiency than white-rot LiPs (lignin peroxidases) [8], suggest a possible contribution of fungal DyPs to lignin biodegradation. Similarly, lignin-degrading capabilities have been claimed for bacterial DyPs [21,22].

Both lignin polymer and substituted anthraquinone dyes, such as RB19 (Reactive Blue 19) (Supplementary Figure S1A), cannot easily access the buried haem cofactor in DyPs and other haem peroxidases. As an alternative for oxidation of these bulky substrates, LRET (long-range electron transfer) from radical-forming aromatic residues at the DyP surface has been suggested [9,23]. Surface residues at the origin of LRET routes were first reported in *Phanerochaete chrysosporium* LiP [24] and *Pleurotus eryngii* VP (versatile peroxidase) [25], and later identified in the sequences of many putative LiPs and VPs from genomes of lignin-degrading white-rot basidiomycetes [20].

Computational analyses can help to explain these LRET processes, requiring, however, the combination of different levels of theory [26]. Long-timescale processes, such as substrate binding, can only be accomplished through MM (molecular mechanics) methods, whereas electron transfer requires QM (quantum mechanics)-based methods, such as QM/MM [27]. The combination of these techniques was shown to be a successful approach in the study of oxidation and electron-transfer processes in haem proteins [28].

In the present study, we expressed *A. auricula-judae* DyP in *E. coli*, solved the crystal structure of the recombinant enzyme and several site-directed variants, and used PELE (Protein Energy Landscape Exploration) [29] to describe the binding of its typical substrate RB19. Subsequent QM/MM analyses of binding sites indicate a preference for substrate oxidation at an exposed tryptophan residue, and identified the LRET pathway to haem. Simultaneously, a mixed tryptophanyl/tyrosyl radical was detected by EPR spectroscopy of the H_2O_2-activated WT (wild-type) DyP. A combined multifrequency EPR and computational approach, together with site-directed mutagenesis studies, enabled the identification of both protein radical contributions. Moreover, we associated a high-turnover site in DyP to the presence of a tryptophanyl radical, in agreement with the QM/MM predictions. In this way, a multidisciplinary evaluation of the role of protein radicals in DyP catalysis is provided.

MATERIALS AND METHODS

Chemicals

Among DyP substrates, RB19, DMP (2,6-dimethoxyphenol), RB5 (Reactive Black 5) and VA (veratryl alcohol) were from Sigma–Aldrich, and ABTS [2,2′-azinobis-(3-ethylbenzothiazoline-6-sulfonic acid)] was from Boehringer Mannheim (see Supplementary Figures S1A–S1E respectively and the Supplementary Methods for other chemicals).

DyP production, activation and purification

The DNA sequence coding mature DyP-I from *A. auricula-judae* (GenBank® accession number JQ650250) [18] was synthesized (ATG:biosynthetics), expressed in *E. coli*, activated *in vitro*, and purified as described in [19] (see the Supplementary Methods for details).

Site-directed mutagenesis and chemical modification of DyP

Simple DyP variants were produced by PCR using the pET23a-DyPI vector harbouring the mature protein-coding sequence of *A. auricula-judae* DyP as a template. For each mutation, direct and reverse primers were designed. For double (or triple) mutations, the mutated vector for the first (or second) mutation was used as template. The pET23a-DyPI plasmids containing the mutations were digested with endonuclease DpnI and transformed into *E. coli* DH5α cells for propagation.

Tryptophan and tyrosine residues in 3 μM WT DyP and the W377S variant were also chemically modified using up to 0.3 mM NBS (*N*-bromosuccinimide) and up to 40 mM TNM (tetranitromethane) (including 2.6% ethanol) respectively [30]. Chemically modified enzymes were used for estimation of residual activity on RB19 (180 μM), DMP (7.5 mM), RB5 (15 μM) and ABTS (1.25 mM) (see the Supplementary Methods

for PCR primers and conditions, and details on chemical modification).

Crystallization, data collection and refinement

Crystallization of WT DyP and five site-directed variants was optimized by the sitting-drop vapour-diffusion method. Crystals of WT DyP were obtained in 32.5% PEG 4000, and those of all the variants were obtained in PEG 2000 MME (30–35%). X-ray diffraction intensities were collected at the SOLEIL (Gyf-sur-Yvette, France) and ALBA (Barcelona, Spain) synchrotrons. The structure of WT DyP and its variants were solved by molecular replacement (see the Supplementary Methods for details; collection, refinement and final statistics are in Supplementary Table S1). Some of the structures did not show electron density for the first two or three residues at the N-terminus, but the whole sequence could be solved for two of them (PDB codes 4W7K and 4W7L). In contrast, the C-terminal region showed good electron density for all of the structures.

Enzyme kinetics

Steady-state kinetic constants were determined from absorbance increases during oxidation of DMP, ABTS and VA at pH 3 (pH 2.5 for VA) measured using a Thermo Spectronic UV–visible spectrophotometer. Absorbance decreases were followed for RB5 and RB19 oxidation (assayed at pH 3 and pH 3.5 respectively) using the same equipment. Eventual changes of enzyme molecular mass after turnover were investigated by MALDI–TOF (see the Supplementary Methods for details). Plotting and analysis of kinetic curves were carried out with SigmaPlot (version 11.0). Apparent affinity, turnover number and catalytic efficiency were estimated by non-linear least-squares fitting to the Michaelis–Menten model. The catalytic efficiency for VA was estimated by linear regression, since no saturation was attained. Calculation of two sets of kinetic constants was performed by adjusting to the Michaelis–Menten model the data from 0.2–10 μM RB19, 4–60 μM DMP, and 0.2–7 μM ABTS, separately from those of 50–270 μM RB19, 200–8000 μM DMP and 30–5000 μM ABTS.

Computational analyses: PELE, MD and QM/MM calculations

The starting structure (based on 4W7J) was prepared at pH 3.5, the optimal pH for RB19 oxidation, by adjusting the protonation state of ionizable residues. Histidine residues were double-protonated, except for His-304 (δ-protonated) and His-115 (ε-protonated), and several aspartic acids (residues 8, 12, 84, 129, 189, 246 and 270) and glutamic acids (residues 158, 220, 225 and 432) were kept in their acidic form. The RB19 atomic charges were derived from QM calculations (see the Supplementary Methods for details of system preparation). Then, RB19 was placed manually in 20 initial random positions on the protein surface and the protein–ligand conformational space was explored with PELE [29]. Results shown are based on 160 independent 48-h PELE simulations. Enhanced local sampling on Trp-377 was obtained with a 5 ns MD simulation allowing us to investigate the effect of solvent and charge fluctuations on the oxidative tendency of Trp-377 and RB19. QM/MM calculations were performed with QSite 5.7 (Schrödinger). Trp-377 LRET pathway calculations were performed with the QM/MM e-pathway approach [32] with His-304–Arg-311, Leu-323–Ala-325, Leu-373–Gln-375 and Asp-395 in the quantum region.
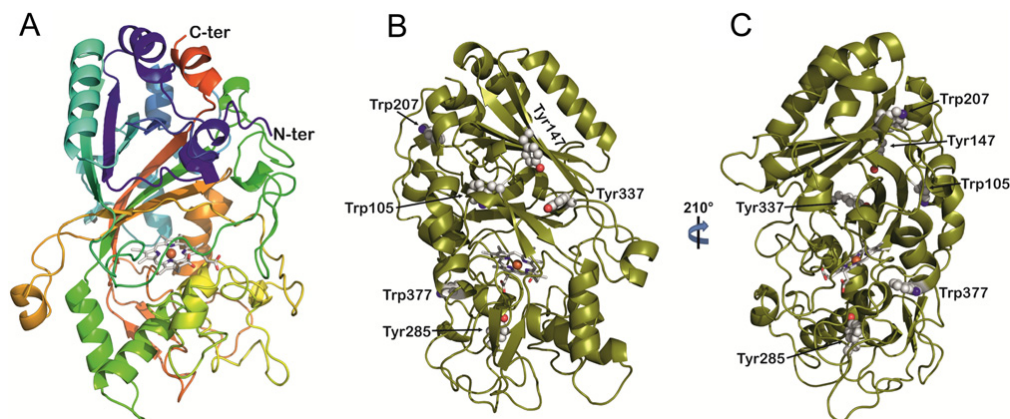
Figure 1   Folding of *A. auricula-judae* DyP and location of exposed aromatic residues

(**A**) General folding constituted by two domains, each of them including two large β-sheets and two or three helices, with the haem cofactor in the upper part of the lower domain [cartoon coloured from the N- to the C-terminus, with the haem shown as CPK (Corey–Pauling–Koltun) sticks]. (**B** and **C**) Location of exposed Trp-105, Tyr-147, Trp-207, Tyr-285, Tyr-337 and Trp-377 (as CPK spheres) in two different orientations of the DyP molecule (cartoon with the haem shown as CPK sticks). From PDB code 4W7J.

EPR spectroscopy and parameter calculations

CW (continuous wave) X-band (9.8 GHz) and W-band (94.17 GHz) experiments were recorded on Bruker Elexsys spectrometers E500 and E600 respectively (see the Supplementary Methods for details). DyP (0.1 μM) activation was carried out using an enzyme/H_2O_2 molar ratio of 1:10 in tartrate, pH 3. H_2O_2 addition was done directly in the EPR tube for the X-band measurements, and the reaction time before freezing was less than 10 s. For the W-band measurements, the H_2O_2 addition was done before filling the EPR tube resulting in a longer freezing time for the sample. Spectra simulations were performed by the Easyspin 4.5.5 package using the 'Pepper' function [33]. Preparatory force field calculations were performed before QM/MM estimation of EPR magnetic parameters. The QM/MM calculations were performed with the MOLCAS 7.4 package [34] coupled with a modified version of the MM package Tinker 4.2. EPR magnetic parameters - g-tensors, hfcc (hyperfine coupling constant) values and Mulliken spin densities - were computed via single-point calculations on the optimized structures using the ORCA2.9 package (F. Neese, University of Bonn, Bonn, Germany). Details of the protocols used to compute the EPR parameters are reported by Bernini et al. [36,37]

RESULTS AND DISCUSSION

Molecular structure: general fold and exposed aromatic residues

The crystal structure of *A. auricula-judae* WT DyP expressed in *E. coli* was solved at 1.79 Å (1 Å = 0.1 nm) resolution (PDB 4W7J), together with those of the Y147S, D168N, W377S, Y147S/W377S and Y147S/G169L/W377S variants (PDB codes 4W7K, 4W7L, 4W7M, 4W7N and 4W7O) solved at 1.05–1.40 Å resolution (Supplementary Table S1). The recombinant DyP is similar (0.48 Å RMSD, 1776 atoms) to the enzyme isolated from a fungal culture (PDB 4AU9). Moreover, most of the variants

show crystal structures largely superimposable with that of WT DyP, except for the mutated residues.

The DyP structure is formed by two domains, each of them including an antiparallel four-stranded large β-sheet and two or three helices resulting in a ferredoxin-like fold (plus two additional β-strands) (Figure 1A). The C-terminal region also includes two small additional helices extending into the N-terminal domain. In spite of the obvious similarity between the two domains, only the C-terminal domain harbours a haem cofactor. His-304 (N_ε) acts as the fifth ligand of the haem iron, with Asp-395 at 2.66 Å. At the opposite side of the haem, Asp-168 and Arg-332 occupy neighbouring positions, suggesting a contribution to the haem reaction with H_2O_2. A single cysteine residue (Cys-299) is present in the C-terminal domain of DyP. The above structural characteristics of the *A. auricula-judae* and other DyPs indicate a common origin with the other members of the CDE superfamily [38] comprising chlorite dismutase [39], DyP and *E. coli* EfeB proteins [13]. Therefore similarities in the haem pocket architecture with the superfamily of classical plant/fungal/prokaryotic peroxidases [40] result from adaptive convergence to provide similar reactivity properties to the haem cofactor (see the Supplementary Results and Discussion).

Near the confluence of the two domains, a channel provides access to the haem cofactor that, due to its location in DyPs, connects to the top of the haem (Supplementary Figure S2). H_2O_2 will enter through this channel to activate the enzyme, forming compound I. However, direct oxidation of typical DyP substrates, such as RB19 and other bulky dyes, by the activated haem is not possible due to the narrow opening of the channel. Therefore LRET appears as a feasible alternative. This is in agreement with the high number of aromatic residues in the *A. auricula-judae* DyP sequence, including seven tyrosine residues and four tryptophan residues. The exposed nature of six of them (Figures 1B and 1C) suggests participation in LRET oxidation by forming reactive radicals at the protein surface. This is reminiscent of that found in ligninolytic peroxidases (LiPs and VPs) where the bulky lignin polymer is oxidized by LRET from an exposed protein radical
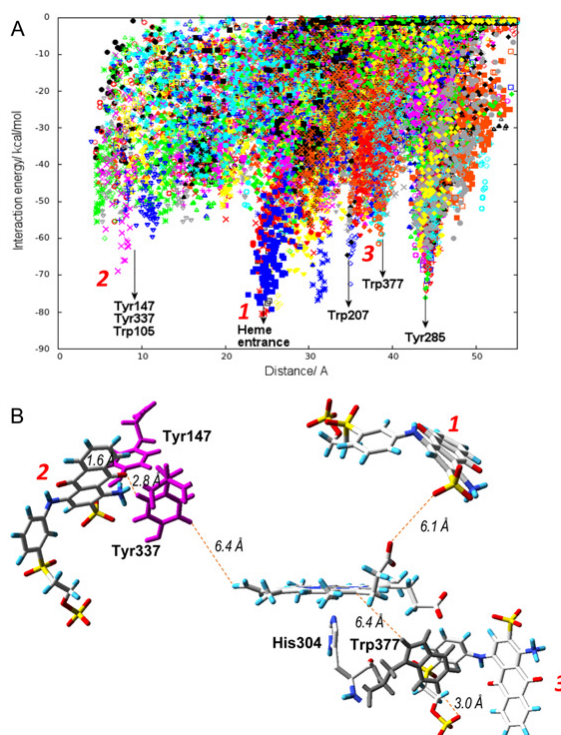
Figure 2   Substrate exploration on the DyP surface

(**A**) Local minima identified in the PELE [29] simulations of RB19 diffusion on the recombinant DyP crystal structure (PDB code 4W7J) showing interaction energy against distance to Tyr-147 (taken as a reference residue). The presence of RB19 in the vicinity of different surface residues and the haem-access channel is indicated. (**B**) Distances between the closest positions of RB19 (magenta sticks) with respect to haem (1), Tyr-147/Tyr-337 (2), and Trp-377 (3) shown by PELE (**A**), and between the above residues and the haem cofactor (the distances are measured including hydrogen atoms). RB19 is shown as CPK (Corey–Pauling–Koltun) sticks, Tyr-147/Tyr-337 as magenta sticks and Trp-377 as grey sticks.

[41–45]. Interestingly, ligninolytic peroxidases have none or only a few tyrosine residues in their sequences, a fact that has been considered as a protection against oxidative inactivation [46]. One remarkable exception is the *Trametopsis cervina* LiP that has a tyrosine residue involved in catalysis [41]. In the molecular models of other LiPs and VPs isolated from fungi or identified from genomes [20], an exposed tryptophan acts as the oxidation site for high-redox-potential aromatics, dyes and polymeric lignin via LRET [47].

Computational simulations: substrate binding (PELE) and LRET pathways (QM/MM)

To identify the possible substrate binding site(s) in *A. auricula-judae* DyP, we performed 160 PELE [29] non-biased simulations, where the typical DyP substrate RB19 was free to explore the structure of the recombinant enzyme. These simulations (Figure 2A) show that RB19 encounters several favourable docking positions on the protein surface, with local minima close to Trp-105, Tyr-147/Tyr-337, Trp-207, Tyr-285 and Trp-377 sites, as well as in the haem channel entrance. Among them, Tyr-147, Tyr337 and Trp-377 are solvent-exposed, but Trp-105 and

Tyr-285 are buried into the protein (Supplementary Figure S3) making their interaction with the dye substrate more difficult. It is important to keep in mind that, whereas simulations were performed with only one substrate molecule (see Supplementary Movie S1 for an example of the exploration), under *in vitro* reaction conditions (with a large excess of substrate) multiple local minima might be populated to different extents. To address computationally which minimum will oxidize the substrate, we used QM/MM studies.

First, we performed a simple QM/MM pairwise comparison between Trp-377 and the other residues identified in the protein exploration with PELE, by including only the two selected residues in the quantum region, subtracting one electron and computing the spin density. There is a clear preference for Trp-377 to be oxidized over Trp-105, Tyr-285 and Tyr-337 (Supplementary Table S2). A comparison of Trp-207 and Trp-377 suggests that both residues could be oxidized. However, electron coupling exponentially decays with donor–acceptor distance and we can therefore exclude Trp-207 due to its large distance from the haem iron (Supplementary Figure S3). To investigate further the oxidation of Trp-377 and Tyr-337 by compound I, we performed new calculations where, in addition to these two residues, the haem was modelled as compound I and included
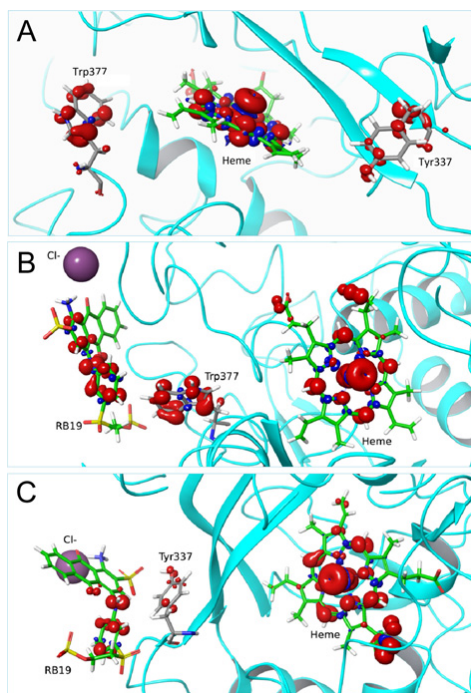
Figure 3   QM/MM electron spin distribution on Trp-377 and Tyr-337 and RB19 substrate

(**A**) Total spin density when including Trp-377, Tyr-337 and haem compound I in the quantum region (in the absence of RB19). (**B** and **C**) Total spin density when, in addition to RB19 and compound I, the quantum region includes Trp-377 or Tyr-337 respectively. From PDB code 4W7J after 5 ns MD (**A**) and selected snapshots from two energy minima showing RB19 near Trp-377 and Tyr-337, during PELE [29] diffusion in Figure 2 (**B** and **C** respectively).

in the quantum region (Figure 3A). The total spin density at Trp-377 shows its preferential oxidation by compound I, validating the previous pairwise analysis, although some density was also observed at Tyr-337.

As a final step, substrate oxidation was investigated in new calculations where we added to the quantum region RB19 at the best PELE position for each of the two residues. Figure 3(B) shows the total spin density for a structure including the dye, the surface Trp-377 and the haem cofactor. Spin density depends on the local electrostatic environment (it was previously found that Trp-377 oxidation was 40 % improved in the presence of a neighbouring Cl^- ion). In agreement with these results, the presence of anionic RB19 enhances Trp-377 oxidation. Similarly, we investigated the possibility of RB19 oxidation in the Tyr-337 site (Figure 3C), where spin density in the tyrosine and substrate are observed. Nevertheless, the most favourable residue for substrate oxidation on the protein surface is Trp-377 and so even though other surface residues may act as potential oxidizing sites, these would have a minor participation in catalysis.

Finally, using QM/MM methods, we were able to map the important residues along the LRET pathway from WT DyP Trp-377 to the haem (Figure 4). This pathway would include a first 3.0 Å electron transfer between the Trp-377 side chain and the
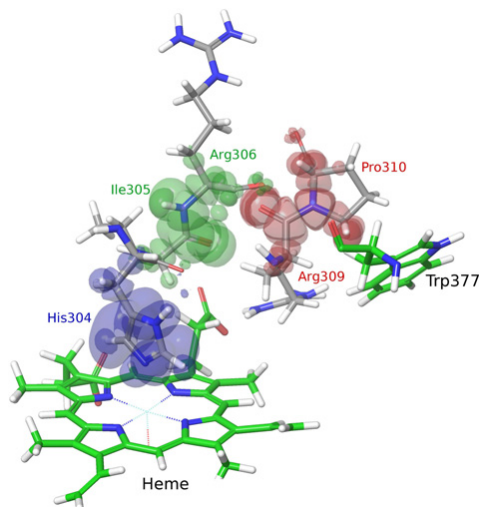

Figure 4   Electron transfer pathway from DyP Trp-377 to haem

The electron transfer pathway was obtained after three iterations of the QM/MM e-pathway approach [32] with a total of 15 residues (His-304–Arg-311, Leu-323–Ala-325, Leu-373–Gln-375 and Asp-395) included in the quantum region. Each iteration identifies the residue(s) with the highest affinity for the electron, and is shown in a different colour. The mapped route includes Pro-310 and Arg-309, followed by Arg-306, Ile-305 and His-304, as shown by the electron spin distribution.

Pro-310/Arg-309 backbone, and a second one (2.9 Å) between the Arg-309 and Arg-306 carbonyls, followed by the Arg-306 to His-304 backbone to reach the haem iron (at only 2.2 Å from the His-304 side chain). The path depicted by the QM/MM e-pathway approach [32] used in the present study is more precise than that previously predicted for the same residue using simpler geometric methods [23].

Catalytic protein radicals: EPR detection in WT DyP and mutated variants

The EPR spectrum of the WT DyP resting state (Figure 5A, top) shows a ferric species prevalently in its axial high spin state ($g_\perp \approx 6$ and $g_\parallel = 2.0$). After adding 10 eq. of H_2O_2, a strong decrease in the ferric signal, and appearance of an intense protein radical signal are evident (Figure 5A, bottom) (the electronic absorption spectra of WT DyP, and the EPR and electronic absorption spectra of the D168N and R332L variants are described in the Supplementary Results and Supplementary Figures S4 and S5 respectively). The yield of the radical observed in the spectrum of the H_2O_2-activated WT DyP is estimated as 0.58 spin/haem. Expansion of the EPR spectrum (Figure 5B) shows a protein radical signal centred at $g = 2.0041(1)$ with two low and high field components, and several sub-splittings. The overall lineshape suggests the presence of two different radical contributions, with the low field side of the spectrum less structured than the high field side, and an intense central line. To identify the radical contributions based on their different anisotropy [48–50], high-frequency EPR spectra were recorded. The narrow scan of the 94 GHz EPR spectrum of H_2O_2-activated WT DyP shows the contributions from two radicals with different g-tensor anisotropy (Figure 5C). For a tryptophanyl radical, 94 GHz EPR (3.3 T) still does not represent the high-field
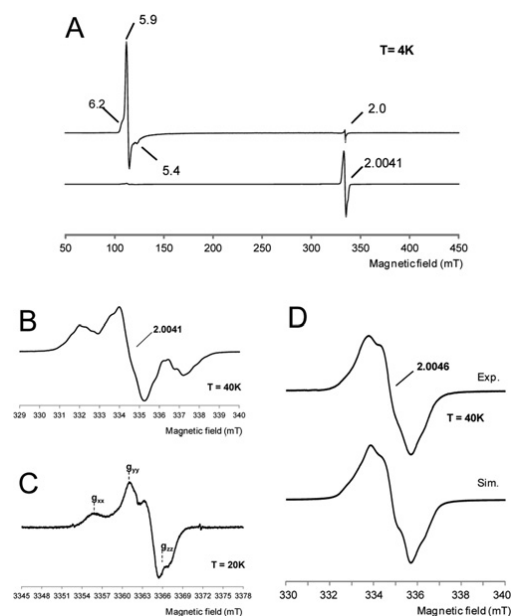
Figure 5   EPR spectra of WT DyP and its W377S variant

(**A**) X-band EPR spectra of WT DyP at pH 3 before (top) and after (bottom) the addition of H$_2$O$_2$ (and rapid freezing). Experimental conditions: $v = 9.39$ GHz, 0.2 mW microwave power, 0.4 mT modulation amplitude. (**B** and **C**) Narrow scan X-band ($v = 9.39$ GHz) and W-band ($v = 94.29$ GHz) respectively of the radical species. The positions of the three g-tensor components of the tyrosyl contribution are indicated. X-band experimental conditions: $v = 9.38$ GHz, 1 mW microwave power and 0.05 mT modulation amplitude; W-band experimental conditions: $v = 94.29$ GHz, 0.05 mW microwave power and 0.1 mT modulation amplitude. (**D**) X-band EPR spectrum of the radical intermediate formed in the W377S variant paired with its better simulation (Sim) (see magnetic parameters in Supplementary Table S3). Experimental conditions: $v = 9.39$ GHz, 1 mW microwave power, 0.2 mT modulation amplitude.

limit where the three g-tensor components are separated, whereas for tyrosyl radicals g-tensor components are well separated at 94 GHz (Figure 5C) and enabled site assignment of the protein radical.

The proof for Trp-377 being involved in the mixed protein radical was provided by the W377S variant. In the high-resolution narrow scan of its 9 GHz EPR spectrum (Figure 5D, top), the tryptophanyl contribution observed for WT DyP (Figure 5B) completely disappeared (the spectrum of the W377S/Y147S variant, not shown, being nearly superimposable). The W377S spectrum (Figure 5D, top) shows a single line with hyperfine resolution at $g = 2.0046(2)$. Simulation of this spectrum at X-band (Figure 5D, bottom) confirmed that it corresponds to a tyrosine phenoxyl radical. This identification was obtained by taking into account the experimental g-tensor component, $g_{xx} = 2.0075$ from the high-field EPR spectrum, and the hf-tensor data from the simulated 9 GHz EPR spectrum. The β-protons' hfccs agree with the computed constants for Tyr-337 (Supplementary Table S3). Therefore this residue would be responsible for the tyrosyl contribution observed in WT DyP, mixed with the main Trp-377 radical contribution.

It had been claimed that Tyr-337 was responsible for substrate oxidation by *A. auricula-judae* DyP based on spin trapping and TNM modification (see below) results [9]. Although their redox

potential is affected by pH and residue environment [51], tyrosyl radicals are less reactive than tryptophanyl radicals, as shown for *P. eryngii* VP whose W164Y variant lost activity on RB5 and VA [42]. In the present study, we confirm that Tyr-337 forms a radical during H$_2$O$_2$ activation of *A. auricula-judae* DyP, but this radical represents a relatively minor contribution of the mixed tryptophanyl/tyrosyl radical signal detected (and Tyr-337 is not catalytically relevant, as discussed below) in agreement with QM/MM spin calculations. Formation of a protein radical was also suggested for *Rhodococcus jostii* DyP [14]. Tryptophanyl and tyrosyl radicals have been identified in different redox enzymes [41,47,52], and it has been suggested that both could be involved in substrate oxidation by DyPs [9,10]. In the present study, we have detected directly, for the first time, a protein radical in a DyP, whose tryptophanyl and tyrosyl contributions were identified by a combined EPR multifrequency and computational approach.

Catalytic properties after chemical modification and site-directed mutagenesis

First, the effect of pH on DyP activity was analysed (Supplementary Figure S6) and the optimal values (pH 3.5 for RB19, pH 2.5 for VA, and pH 3.0 for DMP, ABTS and RB5) were used in subsequent studies. Acidic pH optima were already reported for *A. auricula-judae* DyP [19], and are also typical of lignin-degrading peroxidases (LiP and VP) [46,53]. It is interesting that a delay period was not observed in oxidation reactions with WT DyP, and its W377S variant described below, which showed identical reaction traces and MALDI–TOF molecular masses with/without treatment with VA and H$_2$O$_2$ (Supplementary Figures S7A and S7B). This permits us to rule out in DyP an activation mechanism similar to that of *T. cervina* LiP, which enabled a tyrosine residue to oxidize high-redox-potential substrates after forming a reactive adduct with VA [54].

In a first approach for residue modification, *A. auricula-judae* DyP (3 μM) was treated with NBS, which oxidizes the tryptophan ring to oxindole [55], and TNM, which nitrates the phenolic ring of tyrosine [56]. Near 90 % activity on the four substrates assayed was removed when NBS (up to 300 μM) was used (Figure 6A). This confirms that most DyP activity is associated with tryptophan residue(s), whereas less than 15 % would be due to a different site(s). Similar TNM concentrations did not affect DyP activity, but a partial decrease was observed when a 200-fold higher concentration was used (Figure 6B, white symbols). This indicates that some tyrosine residue(s) contribute (directly or indirectly) to substrate oxidation by DyP, although to a lower extent than the tryptophan residue(s). Additional experiments analysed the effect of chemically modifying tyrosine residues on the very-low-activity W377S variant described below (Figure 6B, black symbols). Although activity decrease with some substrates (such as RB19) was observed, the activity with others (such as ABTS) was only slightly reduced, indicating that, in these cases, residues other than tyrosine residues (or maybe the haem cofactor) are involved.

An interesting initial observation, when analysing the kinetics of DyP oxidations, was the bimodal curves obtained for most substrates (Figure 7A, regions *a* and *b*). Similar sigmoidal kinetic curves were recently reported for DyP [19], and previously for *P. eryngii* VP [57]. Such curves enable calculation of two set of constants (Table 1) and reveal the existence of, at least, two oxidation sites for the same substrate. The *a* and *b* sites in *A. auricula-judae* DyP are characterized by a high turnover (k_{cat}) with low apparent affinity (as shown by the K_m values), and a low turnover with high apparent affinity respectively. The kinetic
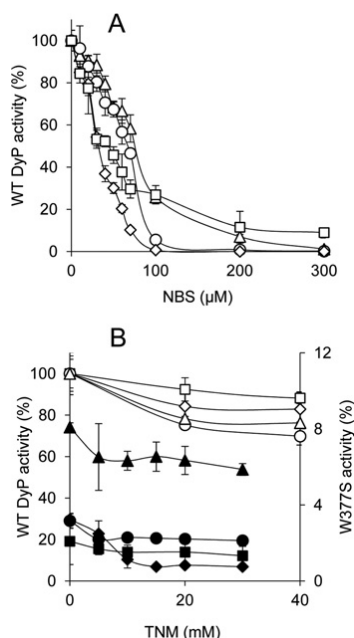
Figure 6   Chemical modification of tryptophan and tyrosine residues in WT DyP and its W377S variant

(**A**) Residual activities of 3 μM WT DyP treated with increasing NBS concentrations (in 50 mM acetate, pH 4) for modification of tryptophan residues. (**B**) Residual activities of 3 μM WT DyP and W377S variant (white and black symbols respectively) treated with increasing TNM concentrations (in 50 mM Tris/HCl, pH 7, with 2.6 % ethanol) for modification of tyrosine residues. The residual activities of the WT DyP and the W377S variant in (**A**) and (**B**) were monitored for oxidation of 180 μM RB19 (diamonds), 7.5 mM DMP (circles), 15 μM RB5 (squares) and 1.25 mM ABTS (triangles), and referred to activities of untreated WT DyP (taken as 100 %).

constants obtained show that the *A. auricula-judae* DyP is very efficient at oxidizing RB19 and ABTS dyes. However, it has lower catalytic efficiencies on DMP and RB5, and extremely low activity on VA. Finally, no Mn^{2+} oxidation activity was observed, as reported for bacterial DyPs [14,15].

Then, the exposed aromatic residues identified in PELE and QM/MM simulations were substituted to verify their eventual involvement in catalysis (information on the effect of mutating haem pocket residues is provided in the Supplementary Results). Aromatic residues were changed to serine in the simple W377S, Y147S and Y337S variants (the latter with low refolding yield). The double variant combining the last two mutations could not be refolded, but Y147F/Y337F could be obtained. Since the haem-access channel was also identified as a possible RB19-binding site (Figure 2A), the G169L variant was also obtained, whose leucine side chain completely blocks the haem-access channel, as shown in the crystal structure of a variant including the G169L mutation, compared with the WT DyP (Supplementary Figures S2B and S2A respectively).

The two sets of constants for oxidation of RB19, DMP and ABTS, and a single set for RB5 and VA, by the above five variants are shown in Table 1. The Y285F variant was also analysed, but no significant modification of the RB19 constants was observed. As a main conclusion, Trp-377 appears responsible

for the high-turnover catalytic site, since its substitution (W377S variant) completely prevented high-turnover oxidation of RB19 and DMP (and decreased over 40-fold the catalytic efficiency of high-turnover ABTS oxidation). In contrast, similar Tyr-337 and Tyr-147 variants (and the double Y147F/Y337F variant) only slightly affected substrate oxidation, and did not remove the high-turnover site.

The above results are illustrated in Figures 7(A)–7(C), where kinetic plots for RB19 oxidation by WT DyP, and several tryptophan (W377S) and tyrosine (Y147S, Y337S and Y147F/Y337F) variants are shown respectively. It is observed that the DyP low-turnover site (Figure 7A, region *b*), remaining after Trp-377 removal (Figure 7B), does not correspond to Tyr-147 or Tyr-337, since changing these residues does not significantly affect the enzyme kinetics (Figure 7C). However, the disappearance of the low-turnover site in the G169L variant (Figure 7D, inset) suggests that this site involves the haem-access channel (although a decrease of the maximal turnover was also observed). Similar results were obtained for ABTS oxidation (Table 1), but the situation could be more complicated for DMP, suggesting that a third, still unidentified, site participates in oxidation of phenols by *A. auricula-judae* DyP. Finally, none of the mutations caused noticeable changes in oxidation of VA, which was highly inefficient in all cases, although Trp-377 appears to be the main residue involved.

A global view of substrate oxidation sites in the *A. auricula-judae* and other basidiomycete DyPs

The enzyme kinetics reveal the existence of at least one high-turnover and one low-turnover oxidation site in *A. auricula-judae* DyP. Site-directed mutagenesis showed that high-turnover oxidation of RB19 and other substrates by DyP takes place at the same tryptophan residue (Trp-377), identified previously as forming a protein radical. This agrees with an 85–100 % decrease in enzyme activity after modifying tryptophan residues with 0.2 mM NBS, and contrasts with the absence of a significant effect reported previously [9]. This discrepancy is probably due to the low NBS concentration used in the latter study. Modification of tyrosine residues with comparatively high TNM concentrations (40 mM) partially decreased (10–25 %) the activity of DyP, as reported previously for the enzyme purified from a fungal culture [23]. However, this does not seem to be due to modification of Tyr-337 (or contiguous Tyr-147) since the enzyme activity was practically unaffected after site-directed mutagenesis of these two residues. A possibility suggested by favourable RB19 docking by PELE is low-turnover oxidation at the DyP haem channel. Interestingly, the *B. adusta* DyP has been crystallized with DMP occupying the entrance of a second channel near the haem propionates [7]. This channel does not exist in the *A. auricula-judae* DyP, where it is covered with a loop (residues 316–324). Therefore direct substrate oxidation by the haem cofactor in the latter DyP is only possible at the main haem access channel, in agreement with results from the G169L mutation that blocked the channel. A tyrosine residue could also be responsible for the second/third catalytic sites, in agreement with the partial activity loss after TNM modification. Nevertheless, once demonstrated that Trp-377 is the main substrate oxidation site in *A. auricula-judae* DyP, it is difficult to unambiguously identify these second/third catalytic sites due to the variety of exposed aromatic residues in this haem peroxidase.

The two aromatic residues forming the mixed trypto-phanyl/tyrosyl radical in the H_2O_2-activated DyP of *A. auricula-judae* (Trp-377 and Tyr-337) are not a peculiarity of this enzyme,
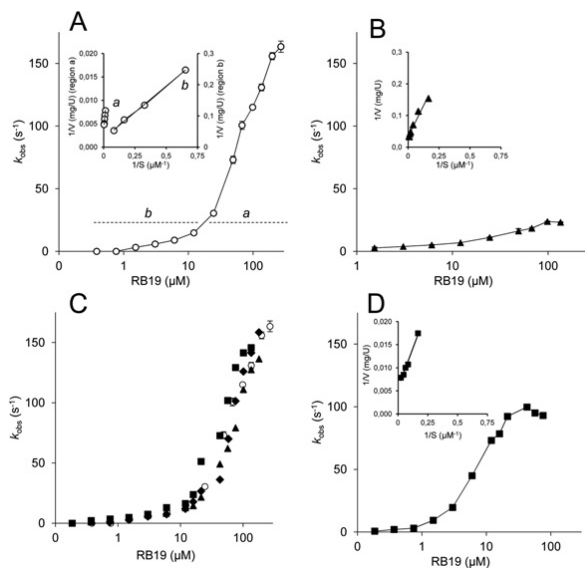
Figure 7   Kinetics for RB19 oxidation by WT DyP and different variants

(**A**) WT DyP biphasic kinetics enabling calculation of two sets of constants in the 0.2–10 μM (*b*) and 50–270 μM (*a*) ranges (inset with Lineweaver–Burk inverse representation for the high, *a*, and low, *b*, turnover sites). (**B**) Simple kinetics yielding a single set of constants (inset, inverse representation) for the W377S variant. (**C**) Results from Y147S (squares), Y337S (diamonds) and Y147F/Y337F (triangles) variants yielding kinetic curves superimposable with that of WT DyP (circles). (**D**) Simple kinetics yielding a single set of constants (inset, inverse representation) for the G169L variant. A substrate concentration logarithmic scale is used in the main plots.

Table 1   Steady-state kinetic constants of WT DyP and five site-directed variants

K_m (μM), k_{cat} (s^{-1}), and k_{cat}/K_m (s$^{-1}\cdot$mM^{-1}) of WT DyP and its Trp-377, Tyr-147, Tyr-337 and Gly-169 variants oxidizing RB19, DMP, ABTS, RB5 and VA (including two sets of constants for the first four substrates corresponding to a high-turnover and a low-turnover site). Means and 95% confidence limits from reactions at 25°C in 0.1 M tartrate, pH 3 (pH 3.5 for RB19 and pH 2.5 for VA) using 0.1 mM H_2O_2 and 10 nM enzyme (100 nM for VA oxidation), are shown (when 1 mM H_2O_2 concentration was used, lower catalytic efficiencies were obtained, although the k_{cat} values often increased). NS, no saturation preventing kinetic constant estimation.

		WT DyP	W377S	Y147S	Y337S	Y147F/Y337F	G169L
RB19 (high turnover)	K_m	90 ± 10	–	95 ± 24	83 ± 3	130 ± 30	10 ± 3
	k_{cat}	224 ± 10	0	175 ± 24	240 ± 47	220 ± 16	106 ± 9
	k_{cat}/K_m	2460 ± 180	–	1860 ± 300	2700 ± 400	1680 ± 170	12400 ± 3000
RB19 (low turnover)	K_m	14.0 ± 2.0	3.9 ± 0.6	6.4 ± 0.8	7.4 ± 2.3	5.8 ± 0.8	–
	k_{cat}	32.0 ± 3.0	8.9 ± 0.6	26.0 ± 6.1	26.0 ± 1.9	18.0 ± 1.9	0
	k_{cat}/K_m	2230 ± 200	2240 ± 240	4070 ± 640	3370 ± 350	3200 ± 200	–
DMP (high turnover)	K_m	703 ± 61	–	763 ± 70	2840 ± 300	4720 ± 460	353 ± 41
	k_{cat}	120 ± 3	0	71 ± 2	228 ± 18	216 ± 7	88 ± 3
	k_{cat}/K_m	200 ± 18	–	93 ± 7	80 ± 1.1	46 ± 3	200 ± 21
DMP (low turnover)	K_m	6.0 ± 0.5	3560 ± 250	38.0 ± 0.4	3.1 ± 0.6	0.7 ± 0.1	6.0 ± 0.4
	k_{cat}	8.0 ± 0.2	6.4 ± 0.1	15.0 ± 1.0	8.4 ± 0.3	3.9 ± 0.1	9.9 ± 0.2
	k_{cat}/K_m	1350 ± 100	1.8 ± 0.1	397 ± 41	2730 ± 520	5900 ± 1000	1600 ± 70
ABTS (high turnover)	K_m	121 ± 7	2750 ± 470	366 ± 30	239 ± 29	173 ± 21	25 ± 2
	k_{cat}	224 ± 3	171 ± 15	311 ± 13	288 ± 10	286 ± 7.2	96 ± 2
	k_{cat}/K_m	1850 ± 94	62 ± 1	850 ± 80	1200 ± 70	1650 ± 180	3770 ± 270
ABTS (low turnover)	K_m	3.1 ± 1.1	30.0 ± 0.5	20.0 ± 6.0	2.0 ± 0.2	17.0 ± 3.6	–
	k_{cat}	7.4 ± 1.4	14.0 ± 1.0	21.0 ± 4.0	2.3 ± 0.3	21.0 ± 6.0	0
	k_{cat}/K_m	2370 ± 420	472 ± 61	1010 ± 100	1100 ± 100	1260 ± 120	–
RB5	K_m	15.6 ± 2.0	10.6 ± 1.0	15.9 ± 1.7	13.4 ± 1.7	39.0 ± 6.9	5.8 ± 0.6
	k_{cat}	4.8 ± 0.2	0.70 ± 0.02	8.3 ± 0.6	5.2 ± 0.3	9.1 ± 2.6	1.2 ± 0.1
	k_{cat}/K_m	310 ± 20	68 ± 4	525 ± 230	400 ± 29	233 ± 20	217 ± 13
VA	K_m	NS	7360 ± 1270	NS	NS	NS	NS
	k_{cat}	–	0.23 ± 0.01	–	–	–	–
	k_{cat}/K_m	0.096 ± 0.002	0.032 ± 0.003	0.101 ± 0.005	0.079 ± 0.011	0.095 ± 0.002	0.130 ± 0.006

but they are significantly conserved among basidiomycete DyPs (Supplementary Figure S8). Among the 65 DyP sequences compared, 58 include a tryptophan residue homologous with Trp-377, and 48 include a tyrosine residue homologous with Tyr-337 (whereas the neighbour Tyr-147 is only conserved in 14 sequences). Therefore the presence of exposed aromatic residues able to form catalytic radicals appears as a common characteristic of basidiomycete, and probably other, DyPs. Although involvement of tyrosyl radicals in oxidation of bulky dyes by DyP has been compared recently with oxidation of recalcitrant aromatics by tryptophanyl radicals in ligninolytic classical peroxidases [4], the results of the present study suggest that tryptophanyl radicals are mainly responsible for substrate oxidation by these two phylogenetically unrelated peroxidase types.

REFERENCES

1 Sugano, Y. (2009) DyP-type peroxidases comprise a novel heme peroxidase family. Cell. Mol. Life Sci. **66**, 1387–1403 CrossRef PubMed
2 Hofrichter, M., Ullrich, R., Pecyna, M.J., Liers, C. and Lundell, T. (2010) New and classic families of secreted fungal heme peroxidases. Appl. Microbiol. Biotechnol. **87**, 871–897 CrossRef PubMed
3 Ruiz-Dueñas, F.J. and Martínez, A.T. (2010) Structural and functional features of peroxidases with a potential as industrial biocatalysts. In Biocatalysts Based on Heme Peroxidases (Torres, E. and Ayala, M., eds), pp. 37–59, Springer-Verlag, Berlin CrossRef
4 Colpa, D.I., Fraaije, M.W. and van Bloois, E. (2014) DyP-type peroxidases: a promising and versatile class of enzymes. J. Ind. Microbiol. Biotechnol. **41**, 1–7 CrossRef PubMed
5 Kim, S.J. and Shoda, M. (1999) Purification and characterization of a novel peroxidase from *Geotrichum candidum* Dec 1 involved in decolorization of dyes. Appl. Environ. Microbiol. **65**, 1029–1035 PubMed
6 Sugano, Y., Muramatsu, R., Ichiyanagi, A., Sato, T. and Shoda, M. (2007) DyP, a unique dye-decolorizing peroxidase, represents a novel heme peroxidase family: Asp^{171} replaces the distal histidine of classical peroxidases. J. Biol. Chem. **282**, 36652–36658 CrossRef PubMed
7 Yoshida, T., Tsuge, H., Hisabori, T. and Sugano, Y. (2012) Crystal structures of dye-decolorizing peroxidase with ascorbic acid and 2,6-dimethoxyphenol. FEBS Lett. **586**, 4351–4356 CrossRef PubMed
8 Liers, C., Pecyna, M.J., Kellner, H., Worrich, A., Zorn, H., Steffen, K.T., Hofrichter, M. and Ullrich, R. (2013) Substrate oxidation by dye-decolorizing peroxidases (DyPs) from wood- and litter-degrading agaricomycetes compared to other fungal and plant heme-peroxidases. Appl. Microbiol. Biotechnol. **87**, 5839–5849 CrossRef
9 Strittmatter, E., Liers, C., Ullrich, R., Wachter, S., Hofrichter, M., Plattner, D.A. and Piontek, K. (2013) First crystal structure of a fungal high-redox potential dye-decolorizing peroxidase: substrate interaction sites and long-range electron transfer. J. Biol. Chem. **288**, 4095–4102 CrossRef PubMed
10 Liers, C., Aranda, E., Strittmatter, E., Piontek, K., Plattner, D.A., Zorn, H., Ullrich, R. and Hofrichter, M. (2014) Phenol oxidation by DyP-type peroxidases in comparison to fungal and plant peroxidases. J. Mol. Catal. B Enzym. **103**, 41–46 CrossRef
11 Zubieta, C., Joseph, R., Krishna, S.S., McMullan, D., Kapoor, M., Axelrod, H.L., Miller, M.D., Abdubek, P., Acosta, C., Astakhova, T. et al. (2007) Identification and structural characterization of heme binding in a novel dye-decolorizing peroxidase, TyrA. Proteins **69**, 234–243 CrossRef PubMed
12 Zubieta, C., Krishna, S.S., Kapoor, M., Kozbial, P., McMullan, D., Axelrod, H.L., Miller, M.D., Abdubek, P., Ambing, E., Astakhova, T. et al. (2007) Crystal structures of two novel dye-decolorizing peroxidases reveal a β-barrel fold with a conserved heme-binding motif. Proteins **69**, 223–233 CrossRef PubMed
13 Liu, X.H., Du, Q., Wang, Z., Zhu, D.Y., Huang, Y., Li, N., Wei, T.D., Xu, S.J. and Gu, L.C. (2011) Crystal structure and biochemical features of EfeB/YcdB from *Escherichia coli* O157: Asp^{235} plays divergent roles in different enzyme-catalyzed processes. J. Biol. Chem. **286**, 14922–14931 CrossRef PubMed
14 Roberts, J.N., Singh, R., Grigg, J.C., Murphy, M. E. P., Bugg, T. D. H. and Eltis, L.D. (2011) Characterization of dye-decolorizing peroxidases from *Rhodococcus jostii* RHA1. Biochemistry **50**, 5108–5119 CrossRef PubMed
15 Brown, M.E., Barros, T. and Chang, M. C. Y. (2012) Identification and characterization of a multifunctional dye peroxidase from a lignin-reactive bacterium. ACS Chem. Biol. **7**, 2074–2081 CrossRef PubMed
16 Sugano, Y., Nakano, R., Sasaki, K. and Shoda, M. (2000) Efficient heterologous expression in *Aspergillus oryzae* of a unique dye decolorizing peroxidase, DyP, of *Geotrichum candidum* Dec 1. Appl. Environ. Microbiol. **66**, 1754–1758 CrossRef PubMed
17 Sugano, Y., Ishii, Y. and Shoda, M. (2004) Role of H164 in a unique dye-decolorizing heme peroxidase DyP. Biochem. Biophys. Res. Commun. **322**, 126–132 CrossRef PubMed
18 Liers, C., Bobeth, C., Pecyna, M., Ullrich, R. and Hofrichter, M. (2010) DyP-like peroxidases of the jelly fungus *Auricularia auricula-judae* oxidize nonphenolic lignin model compounds and high-redox potential dyes. Appl. Microbiol. Biotechnol. **85**, 1869–1879 CrossRef PubMed
19 Linde, D., Coscolín, C., Liers, C., Hofrichter, M., Martínez, A.T. and Ruiz-Dueñas, F.J. (2014) Heterologous expression and physicochemical characterization of a fungal dye-decolorizing peroxidase from *Auricularia auricula-judae*. Protein Expr. Purif. **103**, 28–37 CrossRef PubMed
20 Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R.A., Henrissat, B., Martínez, A.T., Otillar, R., Spatafora, J.W., Yadav, J.S. et al. (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science **336**, 1715–1719 CrossRef PubMed
21 Ahmad, M., Roberts, J.N., Hardiman, E.M., Singh, R., Eltis, L.D. and Bugg, T. D. H. (2011) Identification of DypB from *Rhodococcus jostii* RHA1 as a lignin peroxidase. Biochemistry **50**, 5096–5107 CrossRef PubMed
22 Brown, M.E. and Chang, M.C.Y. (2014) Exploring bacterial lignin degradation. Curr. Opin. Chem. Biol. **19**, 1–7 CrossRef PubMed
23 Strittmatter, E., Wachter, S., Liers, C., Ullrich, R., Hofrichter, M., Plattner, D.A. and Piontek, K. (2013) Radical formation on a conserved tyrosine residue is crucial for DyP activity. Arch. Biochem. Biophys. **537**, 161–167 CrossRef PubMed
24 Doyle, W.A., Blodig, W., Veitch, N.C., Piontek, K. and Smith, A.T. (1998) Two substrate interaction sites in lignin peroxidase revealed by site-directed mutagenesis. Biochemistry **37**, 15097–15105 CrossRef PubMed
25 Pérez-Boada, M., Ruiz-Dueñas, F.J., Pogni, R., Basosi, R., Choinowski, T., Martínez, M.J., Piontek, K. and Martínez, A.T. (2005) Versatile peroxidase oxidation of high redox potential aromatic compounds: site-directed mutagenesis, spectroscopic and crystallographic investigations of three long-range electron transfer pathways. J. Mol. Biol. **354**, 385–402 CrossRef PubMed
26 Wallrapp, F.H., Voityuk, A.A. and Guallar, V. (2013) *In-silico* assessment of protein–protein electron transfer. A case study: cytochrome *c* peroxidase – cytochrome *c*. PLoS Comput. Biol. **9**, e1002990 CrossRef PubMed
27 van der Kamp, M.W. and Mulholland, A.J. (2013) Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. Biochemistry **52**, 2708–2728 CrossRef PubMed

262   D. Linde and others

28 Guallar, V. and Wallrapp, F.H. (2010) QM/MM methods: looking inside heme proteins biochemistry. Biophys. Chem. **149**, 1–11 CrossRef PubMed

29 Borrelli, K.W., Vitalis, A., Alcantara, R. and Guallar, V. (2005) PELE: Protein Energy Landscape Exploration: a novel Monte Carlo based technique. J. Chem. Theory Comput. **1**, 1304–1311 CrossRef

30 Miki, Y., Ichinose, H. and Wariishi, H. (2011) Determination of a catalytic tyrosine in *Trametes cervina* lignin peroxidase with chemical modification techniques. Biotechnol. Lett. **33**, 1423–1427 CrossRef PubMed

31 Reference deleted

32 Guallar, V. and Wallrapp, F. (2008) Mapping protein electron transfer pathways with QM/MM methods. J.R. Soc. Interface **5**, S233–S239 CrossRef PubMed

33 Stoll, S. and Schweiger, A. (2006) EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. J. Magn. Reson. **178**, 42–55 CrossRef PubMed

34 Aquilante, F., De Vico, L., Ferre, N., Ghigo, G., Malmqvist, P.A., Neogrady, P., Pedersen, T.B., Pitonak, M., Reiher, M., Roos, B.O. et al. (2010) Software news and update MOLCAS 7: the next generation. J. Comput. Chem. **31**, 224–247 CrossRef PubMed

35 Reference deleted

36 Bernini, C., Pogni, R., Ruiz-Dueñas, F.J., Martínez, A.T., Basosi, R. and Sinicropi, A. (2011) EPR parameters of amino acid radicals in *P. eryngii* versatile peroxidase and its W164Y variant computed at the QM/MM level. Phys. Chem. Chem. Phys. **13**, 5078–5098 CrossRef PubMed

37 Bernini, C., Andruniow, T., Olivucci, M., Pogni, R., Basosi, R. and Sinicropi, A. (2013) Effects of the protein environment on the spectral properties of tryptophan radicals in *Pseudomonas aeruginosa* azurin. J. Am. Chem. Soc. **135**, 4822–4833 CrossRef PubMed

38 Goblirsch, B., Kurker, R.C., Streit, B.R., Wilmot, C.M. and Dubois, J.L. (2011) Chlorite dismutases, DyPs, and EfeB: 3 microbial heme enzyme families comprise the CDE structural superfamily. J. Mol. Biol. **408**, 379–398 CrossRef PubMed

39 Hofbauer, S., Schaffner, I., Furtmuller, P.G. and Obinger, C. (2014) Chlorite dismutases: a heme enzyme family for use in bioremediation and generation of molecular oxygen. Biotechnol. J. **9**, 461–473 CrossRef PubMed

40 Welinder, K.G. (1992) Superfamily of plant, fungal and bacterial peroxidases. Curr. Opin. Struct. Biol. **2**, 388–393 CrossRef

41 Miki, Y., Calviño, F.R., Pogni, R., Giansanti, S., Ruiz-Dueñas, F.J., Martínez, M.J., Basosi, R., Romero, A. and Martínez, A.T. (2011) Crystallographic, kinetic, and spectroscopic study of the first ligninolytic peroxidase presenting a catalytic tyrosine. J. Biol. Chem. **286**, 15525–15534 CrossRef PubMed

42 Ruiz-Dueñas, F.J., Pogni, R., Morales, M., Giansanti, S., Mate, M.J., Romero, A., Martínez, M.J., Basosi, R. and Martínez, A.T. (2009) Protein radicals in fungal versatile peroxidase: catalytic tryptophan radical in both Compound I and Compound II and studies on W164Y, W164H and W164S variants. J. Biol. Chem. **284**, 7986–7994 CrossRef PubMed

43 Smith, A.T., Doyle, W.A., Dorlet, P. and Ivancich, A. (2009) Spectroscopic evidence for an engineered, catalytically active Trp radical that creates the unique reactivity of lignin peroxidase. Proc. Natl. Acad. Sci. U.S.A. **106**, 16084–16089 CrossRef PubMed

44 Pogni, R., Baratto, M.C., Teutloff, C., Giansanti, S., Ruiz-Dueñas, F.J., Choinowski, T., Piontek, K., Martínez, A.T., Lendzian, F. and Basosi, R. (2006) A tryptophan neutral radical in the oxidized state of versatile peroxidase from *Pleurotus eryngii*: a combined multi-frequency EPR and DFT study. J. Biol. Chem. **281**, 9517–9526 CrossRef PubMed

45 Pogni, R., Baratto, M.C., Giansanti, S., Teutloff, C., Verdín, J., Valderrama, B., Lendzian, F., Lubitz, W., Vázquez-Duhalt, R. and Basosi, R. (2005) Tryptophan-based radical in the catalytic mechanism of versatile peroxidase from *Bjerkandera adusta*. Biochemistry **44**, 4267–4274 CrossRef PubMed

46 Martínez, A.T. (2002) Molecular biology and structure–function of lignin-degrading heme peroxidases. Enzyme Microb. Technol. **30**, 425–444 CrossRef

47 Ruiz-Dueñas, F.J., Morales, M., García, E., Miki, Y., Martínez, M.J. and Martínez, A.T. (2009) Substrate oxidation sites in versatile peroxidase and other basidiomycete peroxidases. J. Exp. Bot. **60**, 441–452 CrossRef PubMed

48 Jeschke, G. (2005) EPR techniques for studying radical enzymes. Biochim. Biophys. Acta **1707**, 91–102 CrossRef PubMed

49 Bleifuss, G., Kolberg, M., Potsch, S., Hofbauer, W., Bittl, R., Lubitz, W., Graslund, A., Lassmann, G. and Lendzian, F. (2001) Tryptophan and tyrosine radicals in ribonucleotide reductase: a comparative high-field EPR study at 94 GHz. Biochemistry **40**, 15362–15368 CrossRef PubMed

50 Svistunenko, D.A., Dunne, J., Fryer, M., Nicholls, P., Reeder, B.J., Wilson, M.T., Bigotti, M.G., Cutruzzola, F. and Cooper, C.E. (2002) Comparative study of tyrosine radicals in hemoglobin and myoglobins treated with hydrogen peroxide. Biophys. J. **83**, 2845–2855 CrossRef PubMed

51 Warren, J.J., Winkler, J.R. and Gray, H.B. (2012) Redox properties of tyrosine and related molecules. FEBS Lett. **586**, 596–602 CrossRef PubMed

52 Stubbe, J. and Der Donk, W.A. (1998) Protein radicals in enzyme catalysis. Chem. Rev. **98**, 705–762 CrossRef PubMed

53 Fernández-Fueyo, E., Ruiz-Dueñas, F.J. and Martínez, A.T. (2014) Engineering a fungal peroxidase that degrades lignin at very acidic pH. Biotechnol. Biofuels **7**, 114 CrossRef

54 Miki, Y., Pogni, R., Acebes, S., Lucas, F., Fernández-Fueyo, E., Baratto, M.C., Fernández, M.I., de los Ríos, V., Ruiz-Dueñas, F.J., Sinicropi, A. et al. (2013) Formation of a tyrosine adduct involved in lignin degradation by *Trametopsis cervina* lignin peroxidase: a novel peroxidase activation mechanism. Biochem. J. **452**, 575–584 CrossRef PubMed

55 Spande, T.F., Green, N.M. and Witkop, B. (1966) Reactivity toward *N*-bromosuccinimide of tryptophan in enzymes zymogens and inhibited enzymes. Biochemistry **5**, 1926–1933 CrossRef PubMed

56 Sokolovsky, M., Riordan, J.F. and Vallee, B.L. (1966) Tetranitromethane: a reagent for nitration of tyrosyl residues in proteins. Biochemistry **5**, 3582–3589 CrossRef PubMed

57 Morales, M., Mate, M.J., Romero, A., Martínez, M.J., Martínez, A.T. and Ruiz-Dueñas, F.J. (2012) Two oxidation sites for low redox-potential substrates: a directed mutagenesis, kinetic and crystallographic study on *Pleurotus eryngii* versatile peroxidase. J. Biol. Chem. **287**, 41053–41067 CrossRef PubMed

SUPPLEMENTARY ONLINE DATA

Catalytic surface radical in dye-decolorizing peroxidase: A computational, spectroscopic and site-directed mutagenesis study

Dolores LINDE[*1], Rebecca POGNI[†1], Marina CAÑELLAS[‡§1], Fátima LUCAS[‡], Victor GUALLAR[‡¶], Maria Camilla BARATTO[†], Adalgisa SINICROPI[†], Verónica SÁEZ-JIMÉNEZ[*], Cristina COSCOLÍN[*], Antonio ROMERO[*], Francisco Javier MEDRANO[*], Francisco J. RUIZ-DUEÑAS[*2], and Angel T. MARTÍNEZ[*2]

[*]Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain, [†]Department of Biotechnology, Chemistry and Pharmacy, University of Siena, I-53100, Siena, Italy, [‡]Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain [§]Anaxomics Biotech, Balmes 89, E-08008 Barcelona, Spain and [¶]ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain

Supplementary methods for DyP production, activation and purification; supplementary results and discussion on electronic absorption and EPR spectra of WT DyP and haem pockets variants; three supplementary tables providing the crystallographic data (**Table S1**), spin density in pairwise comparison of Trp-377 and other residues (**Table S2**) and experimental and computed g-tensors and hfcc values for Tyr-337 radical species (**Table S3**); eight supplementary figures showing the chemical structures of the five DyP substrates assayed (**Fig. S1**), the haem access channel region in WT DyP and after G169L mutation (**Fig. S2**), the location of several tryptophan and tyrosine residues with respect to the haem cofactor (**Fig. S3**), the electronic absorption spectra of WT DyP (**Fig. S4**), the electronic absorption and EPR spectra of the D168N and R332L variants (**Fig. S5**), the optimal pH for DyP oxidation of different substrates (**Fig. S6**), kinetic and MALDI-TOF evidence for lack of a DyP activation mechanism (**Fig. S7**), Tyr-147, Tyr-337 and Trp-377 homolog residues in different clusters of basidiomycete DyP phylogram (**Fig. S8**), and conserved tryptophan and tyrosine residues in several basidiomycete DyPs (**Fig. S9**); one supplementary movie showing a PELE trajectory for RB19 free diffusion on DyP (**Movie S1**); and references for supplementary data are included below.

SUPPLEMENTARY METHODS

Chemicals
DNaseI, polymerase, and NBS were from Boehringer-Mannhein, Roche and Fluka, respectively. Bactopeptone and yeast extract were from Difco. *Nde*I and *Bam*HI were from New England Biolabs. Ampicillin, chloramphenicol, dithiothreitol, hemin, isopropyl-β-D-thiogalactopyranoside, lysozyme, sodium tartrate and TNM were from Sigma-Aldrich. Other chemicals were from Merck.

DyP production, activation and purification
E. coli BL21(DE3)pLysS cells with the pET23a-DyPI vector containing the *A. auricula-judae* mature WT DyP sequence (and those of DyP variants) were grown overnight at 37 ºC and 170 rpm in Luria Bertani broth (with 100 μg/mL of ampicillin and 34 μg/mL of chloramphenicol), and used to inoculate 2 L flasks containing 1 L of Terrific Broth (with ampicillin and chloramphenicol) that were grown for 3 h at 37 ºC and 200 rpm, induced with 1 mM isopropyl-β-D-thiogalactopyranoside, grown for further 4 h, and harvested by centrifugation.

The apoenzyme accumulated in inclusion bodies was solubilised in 50 mM Tris-HCl (pH 8.0) containing 8 M urea, 1 mM EDTA, and 1 mM dithiothreitol for 1 h at 4ºC. Subsequent "in vitro" activation was performed at 4 ºC using 0.2 M urea, 10 µM hemin, 0.02 mM dithiothreitol, 0.1 mM EDTA and 0.1 mg/ml of protein in 50 mM phosphate (pH 6). After 144 h, the folding mixture was concentrated, dialyzed against 20 mM sodium acetate (pH 4.3), and the insoluble material eliminated by centrifugation (13000 rpm, 30 min).

WT DyP and its site-directed variants were purified using a Resource Q column (GE Healthcare) coupled to an ÄKTA liquid chromatography system, using a gradient from 0 to 0.3 M of NaCl in 10 mM Tris-HCl (pH 7). DyPs were analyzed by SDS-PAGE to confirm the purity of the proteins. Absorption spectra were recorded in 10 mM sodium tartrate (pH 5) at 25 ºC in a Thermo Spectronic diode-array spectrophotometer. The DyP molar absorption coefficient (ε_{405} 117000 $M^{-1} \cdot cm^{-1}$) was calculated from (triplicate) Bradford determination of protein concentration, and used to estimate enzyme concentrations. For spectroscopic characterization of the transient states in the DyP catalytic cycle 1-10 eq of H_2O_2 were added to the resting enzyme in 10 mM sodium tartrate (pH 3 and 7).

Site-directed mutagenesis: Primers and PCR conditions

For PCR mutagenesis, direct and reverse primers designed complementary to opposite strands of the DNA region containing the desired mutation were used. The sequences of the direct primers (with the mutated codons in italics) used for mutagenic PCR was the following: Y147S or Y147F mutations 5´-CG TCG ATC TCG AAG CTC (*TCT* or *TTC*, respectively) TCG TTG TCG GCT TCG-3´; D168N mutation, 5´-G TTC GGC TTC CTT *AAC* GGA ATT GCT CAG CC-3´; G169L mutation, 5´-C GGC TTC CTT GAT *CTT* ATT GCT CAG CCC GC-3´; Y285F mutation, 5´-GG AAC AAC AAC TTC ACC *TTC* TCA CAC GCC G- 3´; Y337S or Y337F mutations, 5´-GGC ATC CCC (*TCT* or *TTC*, respectively) GGT CCT GAG GTT ACC-3´; and W377S mutation, 5´-C CTT CAG CAG ACG *TCT* GCC GAT AAC GCC AAC TTC CC-3´.

PCR reactions were carried out in an Eppendorf (Hamburg, Germany) Mastercycler Pro using 10 ng of template DNA, 250 µM each dNTP, 125 ng of direct and reverse primers, 2.5 units of polymerase (Expand Long Template PCR System), and the manufacture´s reaction buffer. Reaction conditions were as follows: i) a "hot start" of 95 ºC for 1 min; ii) 18 cycles at 95 ºC for 50 s, 55 ºC for 50 s, and 68 ºC for 10 min; and iii) a final cycle at 68 ºC for 10 min. The mutated sequences were confirmed by DNA sequencing using an ABI 3730 DNA Analyzer (Applied Biosystem).

Chemical modification

TNM in ethanol (0-40 mM and 0-30 mM for WT DyP and its W377S variant, respectively) or NBS (0-300 µM) were added to 3 µM enzyme in 50 mM Tris-HCl (pH 7) or 50 mM acetate (pH 4), respectively. The final concentration of ethanol in the TNM modification mixture was 2.6%. After 30 min of incubation at 25ºC, the samples were applied into a 30 kDa Centricon Millipore and diafiltrated with 3 vol of 10 mM tartrate (pH 5).

DyP crystallization, data collection and refinement

Crystallization trials with WT DyP and its site-directed variants (Y147S, D168N, W377S,Y147S/W377S and Y147S/G169L/W377S) were carried out by the sitting drop vapor diffusion method, in 96-well MRC2 plates with 50 µl reservoir solution, using the Emerald (Wizard screens I-III) and Jena Biosciences screenings (JBScreen kits 1-10) at 22 ºC. Drops consisted of 0.2 µl of protein solution (10 mg/ml in 10 mM sodium tartrate, pH 5.0) and 0.2 µl of reservoir solution. Crystals of WT DyP were obtained in 32.5% PEG 4000, and crystals of all the site-directed variants were obtained in PEG 2000 MME (30 to 35%).

The crystals were mounted in nylon loops and flash-frozen in liquid N_2 in the mother liquor using Paratone-N (Hampton Research) as cryoprotectant. X-ray diffraction intensities were collected (100 K) at the PROXIMA-1 beam line at SOLEIL (Gyf-sur-Yvette, France), and at XALOC beam line at

ALBA (Barcelona, Spain) using a Pilatus 6M detector. Data were indexed, integrated, merged and scaled using XDS and XSCALE [1].

The structure of DyP was solved by molecular replacement using the crystal structure of *B. adusta* DyP (PDB entry 4AFV) as the search model and the program AUTOMR of the PHENIX package [2]. The final model was obtained by successive refinement rounds (PHENIX package) followed by manual building with Coot [3] using σ_A weighted 2Fo-Fc and Fo-Fc electron density maps. Solvent molecules were introduced in the refinement, as implemented in the PHENIX package, and visually inspected. A total of 5% of reflections was used to calculate the R_{free} value throughout the refinement process. The structures of all the variants were solved using that of WT DyP as a model, and the final models were obtained in the same way as explained above. The structures were validated with MolProbity [4]. PyMOL [5] and Deep view/Swiss Pdb-Viewer (www.expasy.org/spdbv) were used for structure visualization and analysis, and image generation.

Enzyme kinetics

Steady-state kinetic constants of WT DyP and its site-directed variants on five selected substrates (whose structures are shown in **Fig. S1**) were determined using a Thermo Spectronic UV-visible spectrophotometer. For DMP, ABTS and VA oxidation, absorbance increases at 469 nm (dimeric coerulignone ε_{469}= 55,000 cm^{-1} M^{-1}), 436 nm (ABTS cation radical ε_{436}= 29,300 cm^{-1} M^{-1}) and 310 nm (veratraldehyde ε_{310}= 9,300 cm^{-1} M^{-1}) were followed, respectively. Absorbance decreases were followed in the cases of RB5 (ε_{598}= 30,000 cm^{-1} M^{-1}) and RB19 (ε_{595}= 10,000 cm^{-1} M^{-1}) oxidation resulting in dye decolorization. The ability to oxidize Mn^{2+} was investigated by the eventual formation of Mn^{3+}·tartrate complex (ε_{238} 6500 cm^{-1} M^{-1}) in 0.1 M tartrate (pH 5). Reactions were performed at 25 °C in 100 mM tartrate, pH 3, except for VA and RB19 oxidation that pH 2.5 and 3.5 were used, respectively. All the substrates were tested using the enzymes at a final concentration of 10 nM (except VA whose oxidation was assayed with 100 nM enzyme) and 0.1 mM H$_2$O$_2$. Apparent affinity constant (Michaelis–Menten constant, K_m), turnover number (catalytic constant, k_{cat}) and their standard errors were obtained by non-linear least-squares fitting of the experimental measurements to the Michaelis-Menten model. The catalytic efficiency (k_{cat}/K_m) values with their standard errors were calculated fitting the experimental data to the normalized Michaelis-Menten equation: $v = (k_{cat}/K_m)[S]/(1+[S]/K_m)$. The catalytic efficiency constant for VA was obtained by fitting data to a linear regression equation ($k_{cat}= (k_{cat}/K_m)[S]$) since no enzyme saturation was attained.

MALDI-TOF after steady-state turnover

Eventual changes in the molecular mass of WT DyP, similar to those produced during activation of the *T. cervina* LiP by formation of a VA adduct with the catalytic tyrosine residue, were analyzed by MALDI-TOF using an Autoflex III instruments, and 2,5-dihydroxyacetophenone matrix, after enzyme reaction with H$_2$O$_2$ and VA, as previously described [6]. For these analyses, 10 μM enzyme was incubated for 1 h at 25 °C, in sodium tartrate (pH 3) containing 0.5 mM H$_2$O$_2$ and 10 mM VA, and compared with a control without H$_2$O$_2$ and VA. The reaction mixture was centrifuged into a 30 kDa Centricon Millipore, diafiltrated with 3 vol of 20 mM tartrate (pH 5).

Computational analyses: PELE, MD and QM/MM calculations

The model used is based on the recombinant *A. auricula-judae* DyP crystal structure solved here. As the optimum pH for oxidation of RB19 by DyP is 3.5 we prepared the starting structure accordingly. All ionizable residues were inspected with Schrodinger's protein preparation wizard [7] and with the H++ web server [8]. His-304 was assessed to be δ-protonated, His-115 ε-protonated, and the remaining are double protonated (positively charged). At this low pH several aspartic (residues 8, 12, 84, 129, 189, 246 and 270) and glutamic (residues 158, 220, 225 and 432) acids are in their acidic form, while all other ones are found in their anionic states. Electrostatic potential atomic charges on the RB19 substrate (**Fig. S1A**), to be used in PELE and molecular

dynamics (MD), were obtained from an optimization with Jaguar [9] at the DFT/M06-L level with the 6-31G** basis set and a PB implicit solvent.

Once the initial protein structure was prepared, RB19 was placed manually in 20 initial random positions on the protein's surface and the protein-ligand conformational space was explored with PELE [10]. This is a Monte Carlo based algorithm that produces new configurations through a sequential ligand and protein perturbation, side chain prediction and minimization steps. New configurations are then filtered with a Metropolis acceptance test, where the energy is described with an all-atom OPLS force field and a surface generalized Born solvent. In this way it is possible to locate and characterize local and global minima structures for the most favorable protein-ligand interactions. Results shown are based on 160 independent 48-h PELE simulations. Enhanced local sampling on Trp-377 surface site was obtained with a 5 ns MD simulation with DESMOND [11]. Such analysis allowed us to investigate the effect of solvent and charge fluctuations on the oxidative tendency of Trp-377 and RB19.

QM/MM calculations were performed with Qsite [12]. This method allows for the incorporation of the complete protein structure (as well as solvent and ions) at the atomic level while electronic structure based methods are employed in a sub-section of the system. Most calculations were performed at the M06-L(lacvp*)/OPLS level. Spin densities for the highly reactive compound I were computed at the quartet spin state (easier to converge than the isoenergetic doublet one). Spin density on residues was computed by adding all the atomic spin density contributions. Those residues included in the QM region with/without the haem cofactor (compound I) and a substrate (RB19) molecule are indicated in each case. Additionally, electron transfer pathway calculations were performed with the QM/MM e-pathway approach (using the Hartee-Fock method) [13]. This approach maps those residues with higher probability to participate in the electron transfer pathway by finding, iteratively, those regions of the transfer region with lower ionization energy. In particular, we computed the pathway from Trp377 by considering in the QM region His-304, Ile-305, Arg-306, Lys-307, Thr-308, Arg-309, Pro-310, Arg-311, Leu-323, Ser-324, Ala-325, Leu-373, Gln-374, Gln-375 and Asp-395.

EPR equipments

CW X-band (9.8 GHz) measurements were performed with a Bruker E500 Elexsys Series using the Bruker ER 4122SHQE cavity, equipped with an Oxford helium continuous flow cryostat (ESR900). W-band (94.17 GHz) experiments were recorded on a Bruker Elexsys E600 spectrometer, operating in continuous wave, equipped with a 6T split-coils superconducting magnet (Oxford Instrument), using a continuous helium flow cryostat (Oxford Instrument).

SUPPLEMENTARY RESULTS AND DISCUSSION

Haem pocket site-directed variants: Electronic absorption and EPR spectra

Asp-168 in DyP occupies the position of conserved distal histidine in classical peroxidases. This residue and the neighbor Arg-332 are expected to play a central role in catalysis since both the D168N and the R332L variants obtained were fully inactive on the different substrates.

The electronic absorption spectra of DyP resting and H_2O_2-activated (2 eq) states at pH 3 (the optimal pH for DyP activity, as shown in **Fig. S6**) are included in **Fig. S4A**. The latter exhibit the main peak at 403 nm, and small maxima at 529, 558, 597 and 620 nm, all of them being characteristic for compound I. However, the D168N and R332L variants were unable to react with H_2O_2 (pH 3) as shown by their unchanged absorption spectra (**Fig. S5A** and **B**). When 2 eq of H_2O_2 were added to the resting enzyme at pH 7 (which showed reduced absorbance at 406 and 630 nm and the 506 nm maximum displaced to 502 nm with respect to the pH 3 resting state) a compound II-like spectrum, with main peak at 418 nm and small maxima at 530, 554 and 634 nm, was obtained (**Fig. S4B**). Unexpectedly, no compound II-like spectrum could be obtained at pH 3 by self-reduction or using other H_2O_2 concentrations.

The involvement of Asp-168 and Arg-332 in DyP reactivity with H_2O_2, suggested by the electronic absorption spectra, was investigated by EPR. The resting state spectrum of the D168N variant showed a ferric iron in a rhombic environment (g_{xx}= 6.06, g_{yy}=5.6 and g_{zz}= 2.0) (**Fig. S5C**, black trace). After H_2O_2 activation at pH 3 (**Fig. S5C**, red trace) only a small decrease in the ferric iron signal is recorded and no protein radicals are formed. An almost undetectable amount of compound I in the form of the porphyrin radical can be identified superimposed to the ferric g= 2 region of the EPR spectrum (**Fig. S5C**, inset). The results were similar for the R332L variant (**Fig. S5C**), although a small radical signal was observed (4% radical yield for R332L at pH 3 compared with 58% for WT DyP, under the same experimental conditions) (**Fig. S5D**, inset).

The similarities in the haem pocket architecture of unrelated DyP and classical (plant-fungal-prokaryotic) peroxidases would result from adaptive convergence. This is the case of: i) the conserved proximal histidine (*A. auricula-judae* DyP His-304) acting as the fifth haem iron ligand in both enzyme types; and ii) the opposite haem side residues (DyP Asp-168 and Arg-332) most probably contributing to reaction with H_2O_2, as distal histidine and arginine do in classical peroxidases. The latter was demonstrated by the loss of DyP ability to form compound I after the D168N and R332L mutations (as shown by electronic absorption and EPR spectra). A central role of the conserved aspartic acid or arginine in H_2O_2 reaction have been claimed in the *B. adusta* [14] and *R. jostii* DyP [15], respectively, but the present results show that both residues are required for *A. auricula*-judae DyP activity, although some protein radical signal was still observed in the EPR spectrum of the R332L variant. As reported for other DyPs [14], no compound II spectrum was observed during self-reduction of the *A. auricula-judae* DyP (at physiological pH 3), a fact that suggests fastest self-reduction of compound II than of compound I, in contrast with other fungal peroxidases [16]. At the other side of the haem, Asp-395, also conserved in other DyPs [14], would help proximal His-304 to modulate the haem electron-deficiency, as the proximal aspartic acid in classical peroxidases [17].

SUPPLEMENTARY TABLES

Table S1 Data collection and refinement statistics of *A. auricula-judae* WT DyP expressed in *E. coli* and five site-directed variants

	WT DyP	Y147S	D168N	W377S	Y147S/W377S	Y147S/G169L/ W377S
Data collection						
Space group	C2	C2	C2	C2	C2	C2
Cell constants:						
-a (Å)	154.2	184.0	183.7	184.0	154.2	104.4
-b (Å)	100.5	55.9	56.3	56.2	100.5	56.1
-c (Å)	128.6	104.0	103.8	104.1	128.6	82.0
-β (º)	123.7	118.3	117.8	118.0	123.7	96.9
Resolution range (Å)	50.00-1.79 (1.89-1.79)[a]	50.00-1.05 (1.11-1.05)	50.00-1.05 (1.11-1.05)	50.00-1.15 (1.22-1.15)	50.00-1.40 (1.49-1.40)	50.00-1.20 (1.27-1.20)
Total reflections	620794	2174931	2352392	1894778	1142801	949627
Unique reflections	150630	364516	405267	319222	181153	147432
R_{merge} (%)	13.4 (99.5)	8.8 (105.2)	6.8 (50.5)	6.6 (38.4)	7.0 (97.2)	6.0 (45.8)
CC1/2	99.7 (70.5)	99.9 (45.3)	99.9 (82.3)	99.8 (87.3)	99.9 (71.1)	99.9 (93.8)
Completeness (%)	97.1 (90.3)	82.6 (29.8)	91.5 (54.3)	95.2 (72.1)	99.3 (96.8)	99.2 (95.4)
$<I/\sigma(I)>$	6.59 (1.03)	12.3 (1.0)	12.7 (1.7)	13.1 (2.3)	14.2 (1.5)	15.4 (3.1)
Multiplicity	4.1 (3.5)	6.0 (2.9)	5.8 (3.3)	5.9 (3.5)	6.3 (5.2)	6.4 (6.1)
Subunits/asymmetric unit	4	2	2	2	2	1
Wilson B factor (Å^2)	21.52	10.27	8.91	15.35	23.44	12.07
Refinement						
Resolution range	50.00-1.79	50.00-1.05	50.00-1.05	50.00-1.15	50.00-1.79	50.00-1.20
Working reflections	150153	364071	405142	319118	150153	146680
R_{work}/R_{free}	0.193/0.240	0.135/0.157	0.131/0.150	0.131/0.150	0.144/0.184	0.139/0.158
Protein atoms (non H)	13217	6705	6721	6586	6663	3302
Haem group	172	94	94	86	86	43
Water molecules	1564	1582	1887	1156	1489	700
Mean B factors (Å^2):						
-Protein atoms (non H)	24.83	11.86	10.46	17.16	12.65	15.04
-Haem group	19.66	7.90	6.97	10.99	9.27	10.07
-Water molecules	30.29	24.91	23.35	29.60	24.58	27.80
Deviations from ideality:						
-rmsd bond lengths (Å)	0.009	0.009	0.010	0.004	0.009	0.007
-rmsd angles (º)	1.210	1.410	1.432	0.930	1.210	1.257
Ramachandran statistics:						
-Preferred (%)	97.86	98.54	98.66	98.76	98.42	98.87
-Allowed (%)	2.03	1.46	1.23	1.20	1.58	1.13
-Outliers (%)	0.11	0.00	0.11	0.00	0.00	0.00
PDB code	4W7J	4W7K	4W7L	4W7M	4W7N	4W7O

[a]Data in parenthesis correspond to the last resolution layer.

Table S2 Spin density in pairwise QM/MM comparison of Trp-377 and other surface residues

Pairwise comparisons of spin density distribution in Trp-377 and other surface residues at different distances to haem of *A. auricula-judae* DyP were performed by QM/MM calculations (the distances, in parentheses, are between the residue gamma carbon and the haem iron).

Residue 1	Spin (%)	Residue 2	Spin (%)
Trp-377 (12.5 Å)	100	Trp-105 (19.5 Å)	0
Trp-377	38	Trp-207 (32.0 Å)	62
Trp-377	100	Tyr-285 (14.5 Å)	0
Trp-377	100	Tyr-337 (15.9 Å)	0

Table S3 Comparison of experimental (from W377S variant) and computed (for DyP Tyr-337) tyrosine radical EPR parameters

Experimental (W377S variant) and B3LYP/EPR-II g-tensors (g_i) and hfcc values (A_i, in MHz) computed at the B3LYP/AMBER optimized geometries for Tyr-337 radical species are shown.
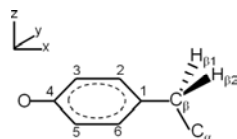
			$^a g_i$	$A_i(H_{\beta 1})$	$A_i(H_{\beta 2})$	$A_i(H_2)$	$A_i(H_6)$	$A_i(H_3)$	$A_i(H_5)$
W377S		x	2.0075	16.6	15.4	2.0	2.7	26.0	24.0
		y	2.0043	17.3	16.0	4.5	5.2	21.0	19.0
		z	2.0020	22.4	21.0	8.1	9.0	7.0	7.0
		Iso	2.0046	18.8	17.5	4.9	5.6	18.0	16.6
B3LYP/EPR-II^b	Tyr-337$^{\cdot}$	x	2.0066	16.6	15.4	2.0	2.7	-26.3	-24.2
		y	2.0044	17.3	16.1	4.5	5.2	-20.3	-18.7
		z	2.0022	22.4	21.5	8.1	9.0	-6.6	-5.7
		Iso	2.0044	18.8	17.7	4.9	5.6	-17.8	-16.2

^{a}The errors on the experimental g values are: $\pm$ 0.0001 for g_x, $\pm$ 0.0002 for g_y and g_z; ^{b}The computed hfcc values for Tyr-337 radical have been used to simulate the experimental EPR spectrum of the radical species in the W377S variant. Although the B3LYP/EPRII g_x value differs from the experimental one, the computed values clearly indicate that the tyrosine radical is embedded in a strongly H-bonded environment, which is in accord with the experiments.
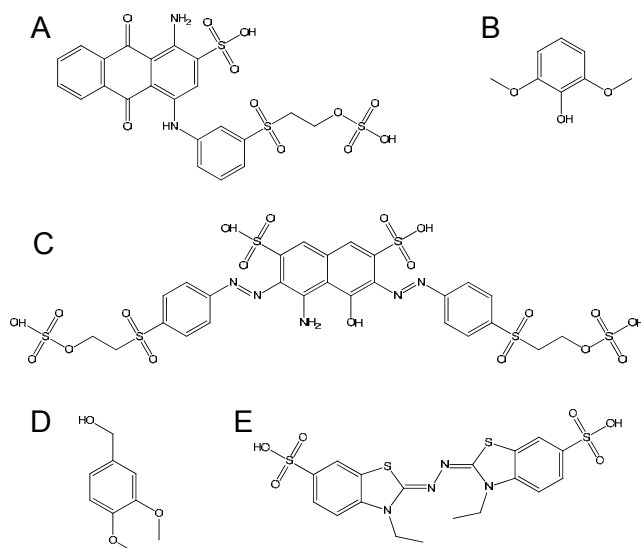
SUPPLEMENTARY FIGURES

Figure S1 Chemical structure of the five DyP substrates assayed
(A) RB19 (1-amino-9,10-dioxo-4-((3-((2-(sulfooxy)ethyl)sulfonyl)phenyl)amino)-9,10-dihydroanthracene-2-sulfonic acid). **(B)** 2,6-Dimeth-oxyphenol. **(C)** RB5 (4-amino-5-hydroxy-3,6-bis((E)-(4-((2-(sulfooxy)ethyl)sulfonyl)phenyl)-diazenyl)-naphthalene-2,7-disulfonic acid). **(D)** Veratryl (3,4-dimethoxybenzyl) alcohol. **(E)** ABTS (2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid).
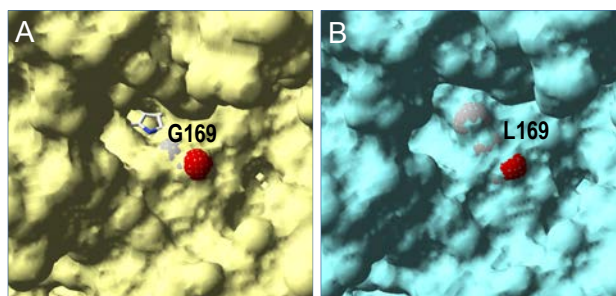

Figure S2 Haem access channel region in WT DyP and after G169L mutation
(A) Solvent access surface in the WT DyP crystal structure (PDB 4W7J) showing channel accessing to the haem cofactor (CPK sticks) and the position of Gly-169 (CPK spheres). **(B)** Solvent access surface in the C169L/Y146S/W377S variant (PDB 4W7O) showing the blocked haem access by the Leu-169 (CPK spheres) side-chain.
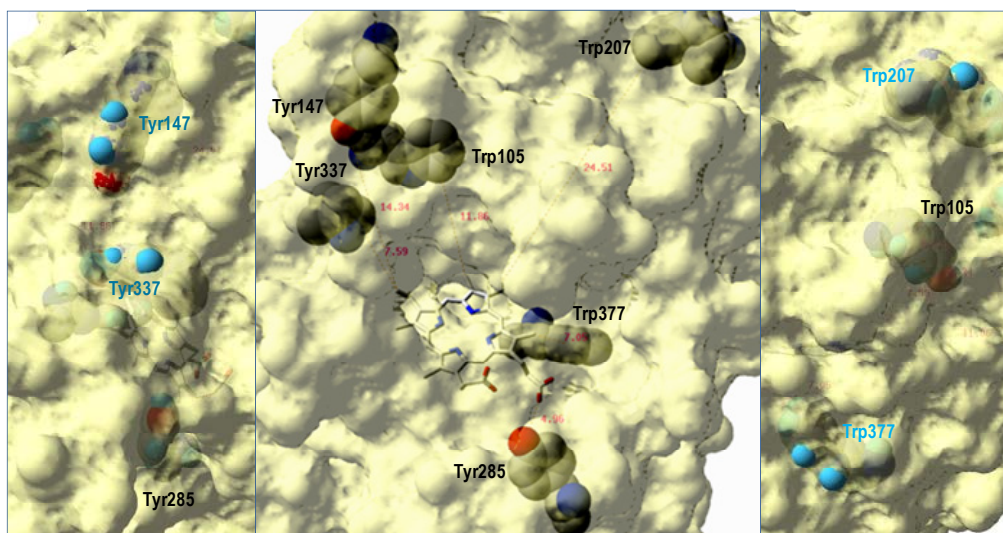
Figure S3 Location of several tryptophan and tyrosine residues near the haem cofactor of DyP.

(center) Semitransparent DyP surface showing the channel accessing the haem cofactor (CPK sticks) and the position of Trp-105, Tyr-147, Trp-207, Tyr-285, Tyr-337 and Trp-377 (CPK spheres) at 12, 14, 25, 5, 8 and 7 Å from the cofactor, respectively. *(left* and *right)* Exposed Tyr-147, Trp-207, Tyr-337 and Trp-377, and buried Trp-105 and Tyr-285 (CPK spheres with H atoms included). From PDB 4W7J.
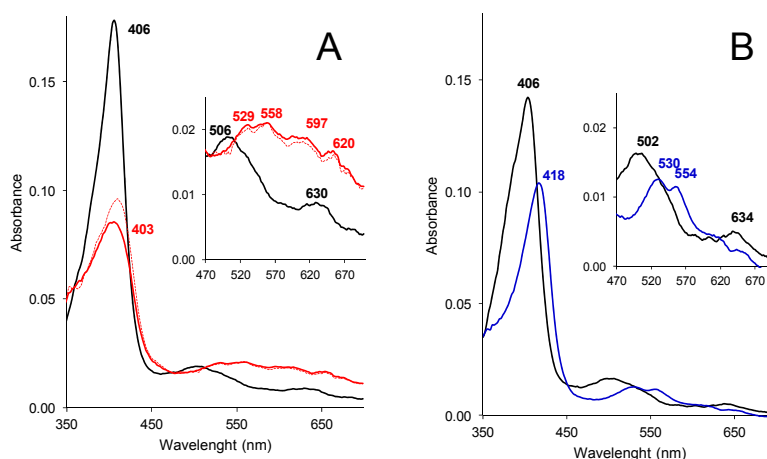

Figure S4 Electronic absorption spectra of DyP

(A) Compound I spectrum (red) obtained 2 s after addition of 2 eq of H_2O_2 to DyP resting state (black) in 10 mM tartrate, pH 3, showing the main Soret band at 403 and 406 nm, respectively, and small charge transfer bands in the inset (dotted line shows the spectrum obtained 0.02 s after H_2O_2 addition). **(B)** Compound II-like spectrum (blue) obtained 2 s after addition of 2 eq of H_2O_2 to DyP resting state (black) in 10 mM Tris-HCl, pH 7, showing the main Soret band at 418 and 406 nm, respectively, and small charge transfer bands in the inset. Spectra were obtained using a stopped-flow rapid spectrophotometer (Bio-Logic) equipped with three-syringe module (SFM300), diode array detector (J&M), and Bio-Kine software.
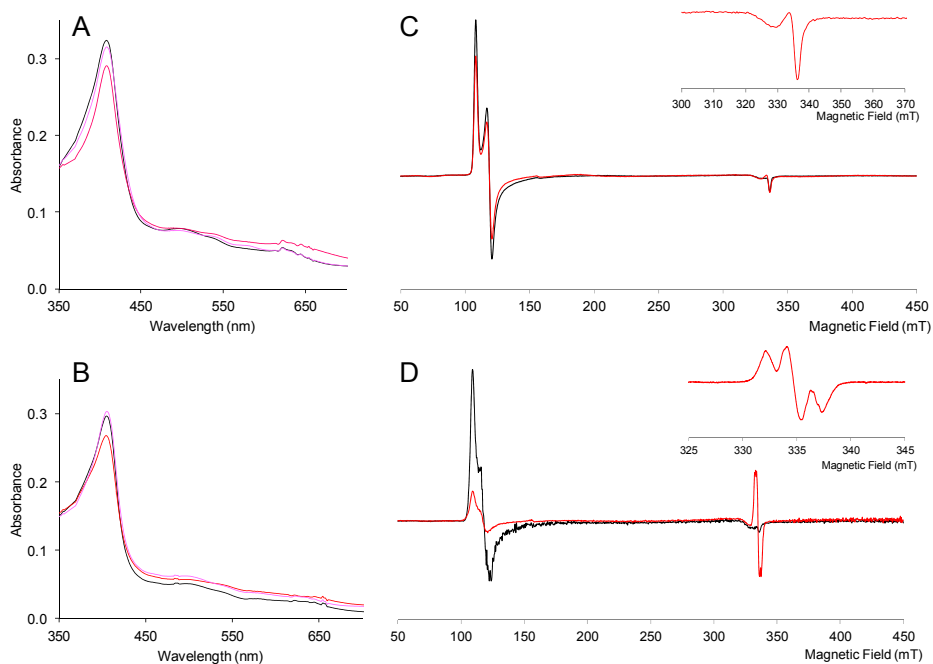
Figure S5 Electronic absorption and EPR spectra of peroxide-activated D168N and R332L variants

(A and **B)** Electronic absorption spectra of 10 μM D168N and R332L variants 0 (black), 50 (red) and 1600 (pink) s after addition of 10 eq of H_2O_2 in 10 mM tartrate, pH 3, respectively. **(C** and **D)** X-band EPR spectra of the D168N and R332L variants (160 μM) before (black) and after (red) addition of 10 eq of H_2O_2 in 70 mM tartrate, pH 3, respectively (insets, detail of the g= 2 region). Experimental conditions: ν= 9.39 GHz, 2 mW microwave power, 0.4 mT modulation amplitude)
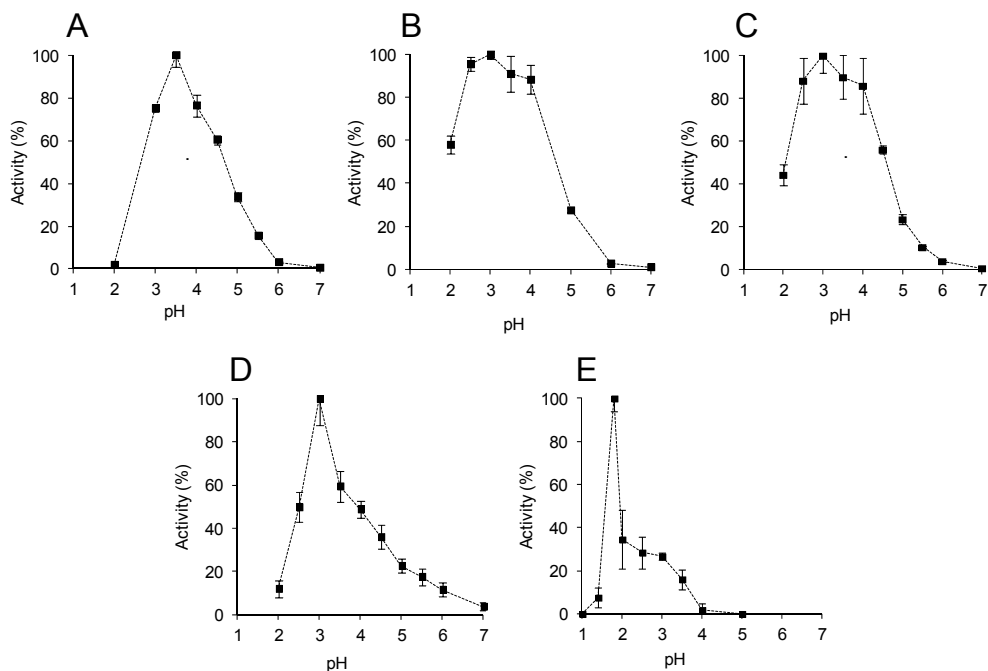
Figure S6 Effect of pH for _A. auricula-judae_ DyP oxidation of different substrates

(A-E) Relative activity of recombinant DyP on RB19 (50 µM), DMP (0.5 mM), ABTS (0.5 mM), RB5 (25 µM) and VA (10 mM), respectively, measured in 50 mM Britton-Robinson buffer (pH 2-12) and KCl-HCl (pH 1) buffers. Bars represent the standard deviations of the means of three measurements
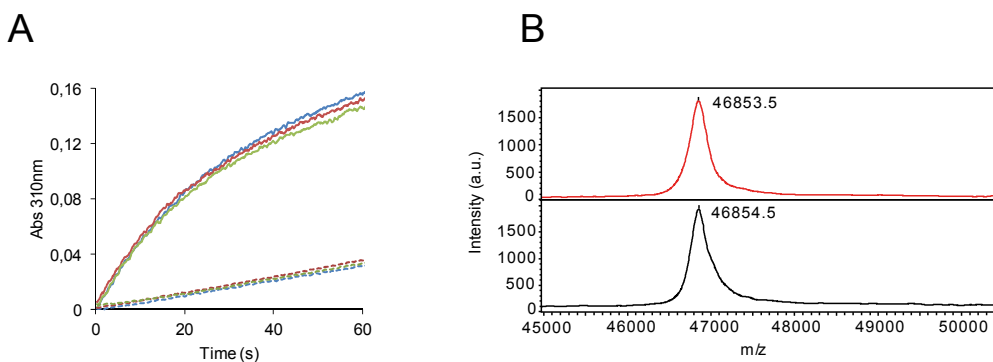

Figure S7 Kinetic and MALDI-TOF evidence for lack of a tyrosine activation mechanism in DyP

(**A**) Absorbance changes at 310 nm (veratraldehyde formation) during oxidation of VA by 10 µM WT DyP (continuous line) and W377S variant (dotted line) previously treated (1 h at 25 ºC) with VA (10 mM) in the presence of 20 µM (red trace) and 100 µM (green trace) H_2O_2, compared with a control incubated with VA alone (blue trace) (reactions in 0.1 M tartrate, pH 2.5, containing 2.5 mM VA, 0.1 mM H_2O_2, and 0.1 mM enzyme). **(B)** MALDI-TOF analysis of the WT DyP (10 µM) treated with 10 mM VA and 0.5 mM H_2O_2 (black profile) compared with an enzyme control incubated without VA and H_2O_2 (red profile). The similar results obtained confirm the absence of a covalent enzyme-VA adduct for activation of a DyP tyrosine residue, as found in _T. cervina_ LiP [6].
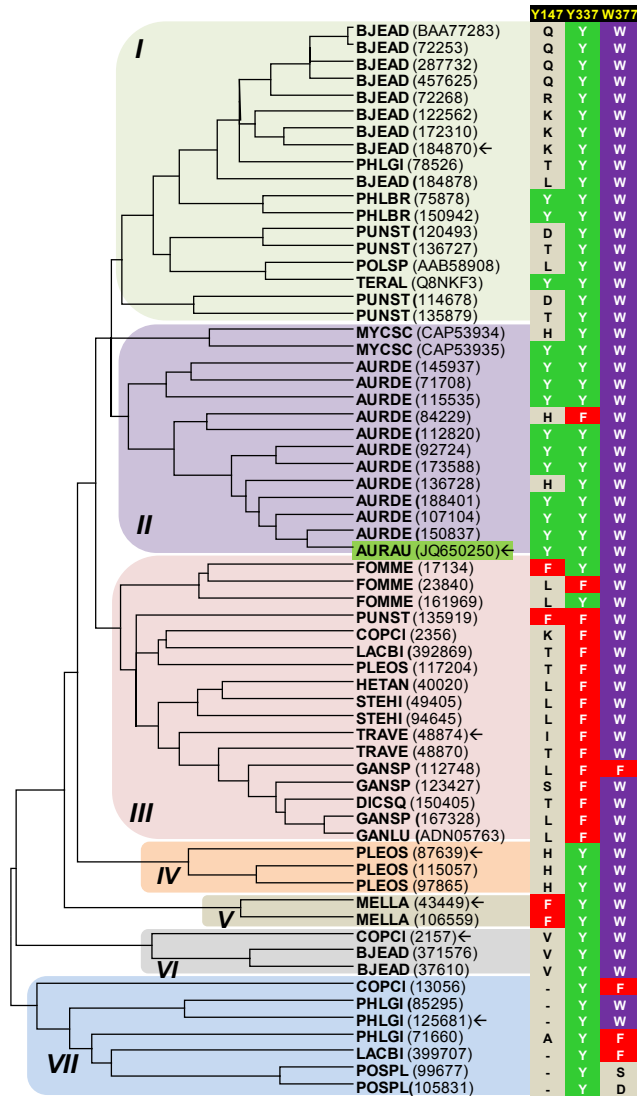
Figure S8 Tyr-147, Tyr-337 and Trp-377 homolog residues in different clusters of basidiomycete DyP phylogram

Sixty-five sequences of basidiomycete DyPs were obtained from the genomes of *Auricularia delicata* (AURAU), *Bjerkandera adusta* (BJEAD), *Coprinopsis cinerea* (COPCI), *Dichomitus squalens* (DICSQ), *Fomitiporia mediterranea* (FOMME), *Ganoderma* sp (GANSP), *Heterobasidion annosum* (HETAN), *Laccaria bicolor* (LACBI), *Melampsora laricis-populina* (MELLA), *Phlebia brevispora* (PHLBR), *Phlebiopsis gigantea* (PHLGI), *Pleurotus ostreatus* (PLEOS), *Postia placenta* (POSPL), *Punctularia strigosozonata* (PUNST), *Stereum hirsutum* (STEHI), and *Trametes versicolor* (TRAVE) available at the JGI (http://genome.jgi.doe.gov) together with GenBank sequences of *A. auricula-judae* (AURAU), *Ganoderma lucidum* (GANLU), *Mycetinis scorodonius* (MYCSC) and *Termitomyces albuminosus* (TERAL). The phylogram obtained with MEGA5 [18] includes three main clusters (I-III and VII; cluster-II containing the *A. auricula-judae* DyP) and three small ones (IV-VI). The JGI and GenBank references are provided, together with indication of the residues occupying the positions homologous to those of *A. auricula-judae* Tyr-147, Tyr-337 and Trp-377 in each of the sequences (details of homology models for the seven DyPs marked with arrows, as representative of the different clusters, are shown in **Fig. S9**).
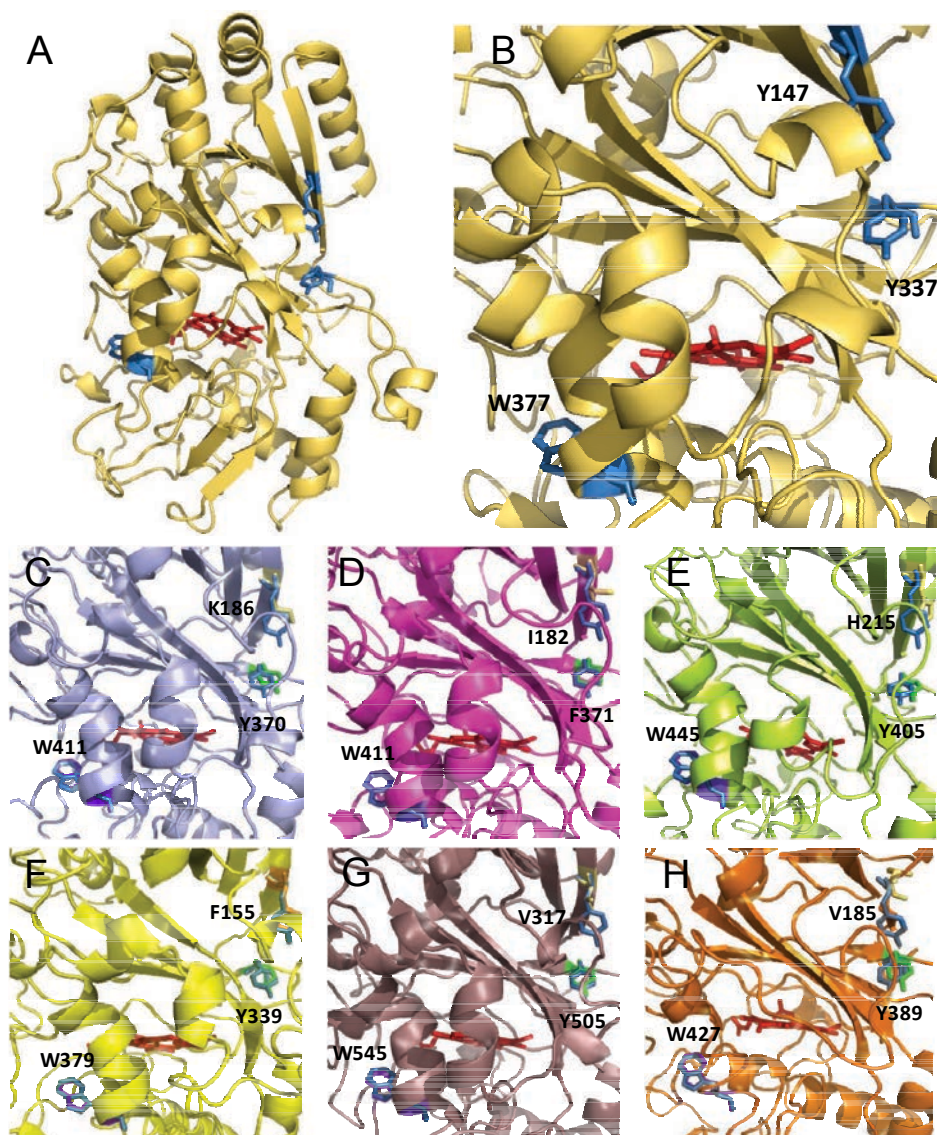
Figure S9 Residues homologous of *A. auricula-judae* DyP Trp147, Tyr337 and Trp377 in basidiomycete DyPs representative for the different clusters of Fig. S8

(A and **B)** Scheme of the whole crystal structure of recombinant *A. auricula-judae* DyP (4W7J) and detail of Tyr-147, Tyr-337 and Trp-377 near the haem cofactor, respectively. **(C-H)** Detail of residues homologous to the above *A. auricula-judae* DyP residues in molecular models for DyPs from the sequenced genomes of *B. adusta* (JGI 184870), *T. versicolor* (JGI 48874), *P. ostreatus* (JGI 87639), *M. laricis-populina* (JGI43449), *C. cinerea* (JGI 2157) and *P. gigantea* (JGI 125681), respectively.

SUPPLEMENTARY MOVIE

Movie S1 PELE trajectory for free substrate diffusion on DyP

One of the 160 free (non-biased) simulations for RB19 diffusion on the DyP molecule provided by PELE software is included, showing substrate binding at the haem access channel and near Trp-377.

REFERENCES FOR SUPPLEMENTARY DATA

1   Kabsch, W. (2010) XDS. Acta Crystallographica Section D-Biological Crystallography, **66,** 125-132
2   Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., Mccoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D-Biological Crystallography, **66,** 213-221
3   Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. D. Biol. Crystallogr., **60,** 2132
4   Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall III, W. B., Snoeyink, J., Richardson, J. S. and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res., **35,** W375-W385
5   de Lano, W. L. (2002) The PyMOLl molecular graphics system, DeLano Scientific, San Carlos, CA, USA
6   Miki, Y., Pogni, R., Acebes, S., Lucas, F., Fernández-Fueyo, E., Baratto, M. C., Fernández, M. I., de los Ríos, V., Ruiz-Dueñas, F. J., Sinicropi, A., Basosi, R., Hammel, K. E., Guallar, V. and Martínez, A. T. (2013) Formation of a tyrosine adduct involved in lignin degradation by *Trametopsis cervina* lignin peroxidase: A novel peroxidase activation mechanism. Biochem. J., **452,** 575-584
7   Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R. and Sherman, W. (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. Journal of Computer-Aided Molecular Design, **27,** 221-234
8   Anandakrishnan, R., Aguilar, B. and Onufriev, A. V. (2012) H++3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Res., **40,** W537-W541
9   Schrödinger. (2011) Jaguar 7.8, LCC, New York
10 Borrelli, K. W., Vitalis, A., Alcantara, R. and Guallar, V. (2005) PELE: Protein energy landscape exploration. A novel Monte Carlo based technique. J. Chem. Theory Comput., **1,** 1304-1311
11 Bowers, K. J., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossváry, I., Moraes, M. A., Sacerdoti, F. D., Salmon, J. K., Shan, Y. and Shaw, D. E. (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. Proc. ACM/IEEE Conference on Supercomputing (SC06), Tampa, 11-17 November, 43
12 Schrödinger. (2011) QSite 5.7, LCC, New York
13 Guallar, V. and Wallrapp, F. (2008) Mapping protein electron transfer pathways with QM/MM methods. J. R. Soc. Interface, **5,** S233-S239
14 Sugano, Y., Muramatsu, R., Ichiyanagi, A., Sato, T. and Shoda, M. (2007) DyP, a unique dye-decolorizing peroxidase, represents a novel heme peroxidase family. Asp171 replaces the distal histidine of classical peroxidases. J. Biol. Chem., **282,** 36652-36658
15 Singh, R., Grigg, J. C., Armstrong, Z., Murphy, M. E. P. and Eltis, L. D. (2012) Distal heme pocket residues of B-type dye-decolorizing peroxidase: Arginine but not aspartate is essential for peroxidase activity. J. Biol. Chem., **287,** 10623-10630
16 Pérez-Boada, M., Ruiz-Dueñas, F. J., Pogni, R., Basosi, R., Choinowski, T., Martínez, M. J., Piontek, K. and Martínez, A. T. (2005) Versatile peroxidase oxidation of high redox potential aromatic compounds: Site-directed mutagenesis, spectroscopic and crystallographic investigations of three long-range electron transfer pathways. J. Mol. Biol., **354,** 385-402
17 Martínez, A. T. (2002) Molecular biology and structure-function of lignin-degrading heme peroxidases. Enzyme Microb. Technol., **30,** 425-444
18 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol., **28,** 2731-2739

3.1.2 *Article II:* **Steroid hydroxylation by basidiomycete peroxygenases: a combined experimental and computational study.**

Esteban D. Babot,[a] José C. del Río,[a] Marina Cañellas,[b,c] Ferran Sancho,[b,c] Fátima Lucas,[b] Victor Guallar,[b,d] Lisbeth Kalum,[e] Henrik Lund,[e] Glenn Gröbe,[f] Katrin Scheibner,[f] René Ullrich,[g] Martin Hofrichter,[g] Angel T. Martínez,[h] Ana Gutiérrez[a]

[a] Instituto de Recursos Naturales y Agrobiología de Sevilla, CSIC, Seville, Spain

[b] Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Barcelona, Spain

[c] Anaxomics Biotech, Barcelona, Spain

[d] ICREA, Barcelona, Spain

[e] Novozymes A/S, Bagsvaerd, Denmark

[f] JenaBios GmbH, Jena, Germany

[g] TU Dresden, Department of Bio- and Environmental Sciences, Zittau, Germany

[h] Centro de Investigaciones Biológicas, CSIC, Madrid, Spain

Address correspondence to Ana Guitérrez, anagu@irnase.csic.es

Steroid hydroxylation by basidiomycete peroxygenases: A combined experimental and computational study

Esteban D. Babot,[a] José C. del Río,[a] Marina Cañellas,[b,c] Ferran Sancho,[b,c] Fátima Lucas,[b] Víctor Guallar,[b,d] Lisbeth Kalum,[e] Henrik Lund,[e] Glenn Gröbe,[f] Katrin Scheibner,[f] René Ullrich,[g] Martin Hofrichter,[g] Angel T. Martínez[h] and Ana Gutiérrez[a*]

[a]Instituto de Recursos Naturales y Agrobiología de Sevilla, CSIC, Reina Mercedes 10, E-41012 Seville, Spain [b]Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain

[c]Anaxomics Biotech, Balmes 89, E-08008 Barcelona, Spain [d]ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain [e]Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark [f]JenaBios GmbH, Orlaweg 2, 00743 Jena, Germany

[g]TU Dresden, Department of Bio- and Environmental Sciences, Markt 23, 02763 Zittau, Germany

[h]Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain

* Address correspondence to Ana Gutiérrez, anagu@irnase.csic.es

ABSTRACT

The goal of this study is the selective oxyfunctionalization of steroids under mild and environmentally-friendly conditions using fungal enzymes. With this purpose, peroxygenases from three basidiomycete species were tested for hydroxylation of a variety of steroidal compounds, using H_2O_2 as the only co-substrate. Two of them are wild-type enzymes from *Agrocybe aegerita* and *Marasmius rotula*, and the third one is a recombinant enzyme from *Coprinopsis cinerea*. The enzymatic reactions on free and esterified sterols, and steroid hydrocarbons and ketones were followed by gas chromatography, and the products were identified by mass

spectrometry. Hydroxylation at the side chain over the steroidal rings was pre-
ferred, with the 25- hydroxyderivatives predominating (interestingly antiviral and
other biological activities of 25-hydroxycholesterol have been recently reported).
However, hydroxylation in the ring moiety and terminal hydroxylation at the side-
chain was also observed in some steroids, the former favored by the absence of
oxygenated groups at C3 and by the presence of conjugated double bonds in the
rings. To understand the yield and selectivity differences between the different ster-
oids, a computational study was performed using Protein Energy Landscape Ex-
ploration (PELE) software for dynamic ligand diffusion. These simulations showed
that the active site geometry and hydrophobicity favors the entrance of the steroid
side-chain, while the entrance of the ring is energetically penalized. Also, a direct
correlation between the conversion rate and the side-chain entrance ratio could be
established, that explains the varying reaction yields observed.

KEYWORDS: basidiomycete peroxygenases, steroids, selective hydroxylation,
GC-MS, PELE simulation

Steroids represent an important class of natural compounds that are wide-
spread in nature and have a multitude of pharmacological properties. Indeed, ster-
oids are ranked among the most marketed medical products and represent the sec-
ond largest category next to antibiotics. The physiological activity of steroids de-
pends on their structure, including the oxidation state of the rings, and the type,
number and regio and stereo position of the functional groups attached (1). It is
known that even minor changes in the structure of steroids can highly affect their
biological activity, which has promoted countless studies on the modification of
naturally occurring steroids in search of new and more active compounds (2).
Among these modifications, hydroxylation is one of the most important reactions
in steroid oxyfunctionalization.

Hydroxylation serves to increase the polarity of the rather hydrophobic
steroids, and hydroxylated steroids often express a higher biological activity as
compared with their less polar non-hydroxylated analogs. Moreover, hydroxylation
can be used to develop intermediates for further synthesis, by offering access to
otherwise inaccessible sites of the steroid molecule. Mostly, these modifications
have been obtained applying microorganisms (1) with only few examples where
isolated enzymes were used. Members of the cytochrome P450 monooxygenase
(P450) superfamily are notable examples of such enzymatic catalysts, since be-
cause of their catalytic versatility, would perfectly meet the requirements of chem-
ical synthesis (3, 4). However, large-scale applications are not yet feasible due to
the intrinsic properties of these oxygenases (5), whose activation requires

NAD(P)H as electron donor and auxiliary flavin-reductases (or a second flavin domain) for electron transfer to O_2 (6).

A few years ago, a new peroxidase type was discovered in the wood-dwelling basidiomycete *Agrocybe aegerita* (7), which turned out to be a true peroxygenase efficiently transferring oxygen from peroxide to various organic substrates (8). This peroxygenase is able to catalyze reactions formerly assigned only to P450s (9). However, unlike P450s that are intracellular enzymes whose activation often requires an auxiliary enzyme and a source of reducing power, the *A. aegerita* enzyme is a secreted protein, therefore far more stable, and only requires H_2O_2 for activation (8). The *A. aegerita* peroxygenase has been shown to catalyze numerous interesting oxygenation reactions on aromatic compounds (10) and recently, the action of this enzyme on aliphatic compounds was demonstrated expanding its biotechnological interest (11, 12). The enzyme, first described as a haloperoxidase and later mostly referred to as aromatic peroxygenase (APO), is now named "unspecific peroxygenase" (UPO; EC 1.11.2.1).

After the first peroxygenase of *A. aegerita* (7), similar enzymes have also been found in other basidiomycetes, such as *Coprinellus radians* (13) and *Marasmius rotula* (14) and there are indications for their widespread occurrence in the fungal kingdom (15). Recently, over one-hundred peroxygenase-type genes (encoding enzymes of the heme-thiolate peroxidase superfamily) have been identified during the analysis of 24 basidiomycete genomes (16) including *Coprinopsis cinerea* (17). The wild-type *C. cinerea* peroxygenase has not been isolated to date from this fungus, but one of the peroxygenase genes from its genome (protein model 7249 available at the JGI; http://genome.jgi.doe.gov/Copci1) was heterologously expressed by Novozymes A/S (Bagsvaerd, Denmark). This first recombinant peroxygenase can be a powerful biocatalyst for synthetic applications because of the potential high expression yield, and the possibility to tune its catalytic and operational properties using protein engineering tools. Recently, the ability of this recombinant peroxygenase in the oxyfunctionalization of several aliphatic compounds was demonstrated (18). Here, the efficiency of three different peroxygenases, namely the wild enzymes from *A. aegerita* (*Aae*UPO) and *M. rotula* (*Mro*UPO) and the recombinant enzyme from *C. cinerea* (r*Cci*UPO), in oxyfunctionalization of a variety of steroidal compounds (including free and esterified sterols, and steroid hydrocarbons and ketones) is evaluated for the first time. All three fungi are agaric basidiomycetes, but they belong to different families of the order Agaricales, and they also differ to some extent from an eco-physiological point of view. *M. rotula* (family Marasmiaceae) is a fungus that stands between white-rotters and litter- decomposers because it preferably colonizes small twigs of broad-

leaved trees (*Fagus*, *Quercus*, *Acer*). *C. cinerea* (family Psathyrellaceae) naturally dwells on older dung and in soils rich in organic nitrogen (i.e. it is a specialized litter decomposer). *A. aegerita* (family Bolbitaceae) is a wood-dwelling fungus that causes an unspecific white-rot on truncs and stumps of broad-leaved trees (e.g. *Populus* and *Acer* species). However, it would be daring to conclude from these evolutionary and ecological differences that the UPOs of these fungi must be different.

In addition to the experimental assays, where the products from the reaction of a variety of steroids with the three above peroxygenases were analyzed by gas-chromatography-mass spectrometry (GC-MS), a set of representative molecules was selected for computational simulations with *Aae*UPO (for which a crystal structure is available) to get further insights into the molecular determinants that affect the reactivity of the different steroids types. All simulations followed an identical protocol where, after appropriate preparation of the protein with heme as compound I (the two-electron oxidized cofactor containing a Fe(IV)=O porphyrin cation radical complex), each substrate was placed at the entrance of the heme access channel, still in the solvent. From there the substrate explored both the entrance pathway and the active site using the Protein Energy Landscape Exploration (PELE) software (19, 20). From the information provided by the energy profiles and trajectories analyses the effect of the structural differences between the substrates and their peroxygenase reactivity could be rationalized.

MATERIALS AND METHODS

Enzymes. r*Cci*UPO was provided by Novozymes A/S (Bagsvaerd, Denmark). This recombinant enzyme corresponds to the protein model 7249 from the sequenced *C. cinerea* genome available at the JGI (http://genome.jgi.doe.gov/Copci1) expressed in *Aspergillus oryzae* (patent WO/2008/119780). The protein was purified using a combination of S- Sepharose and SP-Sepharose ion-exchange chromatography. The recombinant peroxygenase preparation is an electrophoretically homogeneous glycoprotein with a molecular mass around 44 kDa (a non-uniform glycosylation pattern was observed), a typical UV-vis spectrum with a Soret band at 418 nm, and the ability to oxygenate different aromatic compounds with a specific activity of approximately 100 U·mg^{-1} (measured as described below). The *Aae*UPO and *Mro*UPO enzymes are two wild-type peroxygenases isolated from cultures of *A. aegerita* DSM 22459 and *M. rotula* DSM 25031, deposited at the German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany). *Aae*UPO was produced in suspensions of soybean meal and purified by several steps of fast protein liquid chromatography (FPLC) using different ion exchangers (SP-Sepharose,

MonoQ and MonoS) with a size exclusion chromatography (Superdex 75) as final isolation step. The final protein fraction had a molecular mass of 46 kDa with one enriched isoform (*Aae*UPO II). This isoform was used in the present study (7, 21). The *Mro*UPO was purified by FPLC to apparent homogeneity, confirmed by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) under denaturing conditions, and showed a molecular mass of 32 kDa and isoelectric point of pH 5.0-5.3. The UV-visible spectra of the enzymes showed a characteristic maximum around 420 nm (Soret band of heme-thiolate proteins) (14). All media and columns used for enzyme isolation were purchased from GE Healthcare Life Sciences. One UPO activity unit is defined as the amount of enzyme oxidizing 1 μmol of veratryl alcohol to veratraldehyde (ε_{310} 9300 M^{-1}·cm^{-1}) in 1 min at 24ºC, pH 7 (the optimum for peroxygenase activity) after addition of H_2O_2 (0.5 mM H_2O_2 in r*Cci*UPO reactions, and 2.5 mM H_2O_2 in *Aae*UPO and *Mro*UPO reactions).

Steroids. The model steroid compounds (from Sigma-Aldrich) used in the enzymatic reactions include: i) free sterols, such as cholesterol, campesterol, ergosterol, sitosterol and stigmasterol; ii) steroid ketones, such as cholestan-3-one, 4-cholesten-3-one, cholesta-3,5- dien-7-one and testosterone; iii) steroid hydrocarbons, such as cholestane, cholesta-3,5-diene and pregnane; and iv) sterol esters, such as cholesteryl acetate, cholesteryl butyrate and cholesteryl caprylate.

Enzymatic reactions. Reactions of the model steroids (0.05 mM concentration) with the three peroxygenases (1 U) were performed in 5-mL vials containing 50 mM sodium phosphate (pH 7 in *Aae*UPO and r*Cci*UPO reactions, and pH 5.5 in *Mro*UPO reactions) at 40ºC and 60 min reaction time, in the presence of H_2O_2 (0.5 mM in r*Cci*UPO reactions, and 2.5 mM in *Aae*UPO and *Mro*UPO reactions). Prior to use, the substrates were dissolved in acetone and added to the buffer to give a final acetone concentration of 40% (vol/vol) in most cases. In control experiments, substrates were treated under the same conditions (including 0.5 mM and 2.5 mM H_2O_2) but without enzyme. Products were recovered by liquid-liquid extraction with methyl *tert*-butyl ether and dried under N_2. *N,O*- Bis(trimethylsilyl)trifluoroacetamide (Supelco) was used to prepare trimethylsilyl (TMS) derivatives that were analyzed by GC-MS.

GC-MS analyses. The analyses were performed with a Varian 3800 chromatograph coupled to an ion-trap detector (Varian 4000) using a medium-length fused-silica DB-5HT capillary column (12 m x 0.25 mm internal diameter, 0.1 μm film thickness) from J&W Scientific (22). The oven was heated from 120°C (1 min) to 300°C (15 min) at 10°C·min^{-1}. The injector was programmed from 60°C (0.1 min) to 300°C (28 min) at 200°C·min^{-1}. The transfer line was kept at 300°C and helium

was used as carrier gas (2 ml·min^{-1}). For some analyses a Shimadzu GC-MS QP2010 Ultra, with a fused-silica DB-5HT capillary column (30 m x 0.25 mm internal diameter, 0.1 μm film thickness) was also used. The oven was heated from 120°C (1 min) to 300°C (15 min) at 5°C·min^{-1}. The injection was performed at 300°C and the transfer line was kept at 300°C. Compounds were identified by mass fragmentography and comparing their mass spectra with those of the Wiley and NIST libraries and standards. Quantification was obtained from total-ion peak area, using response factors of the same or similar compounds. Data from replicates were averaged and in all cases (substrate conversion and relative abundance of reaction products) the standard deviations were below 3.5% of the mean values.

PELE computational analyses. For the computational study eight representative compounds (cholesterol, sitosterol, cholestan-3-one, cholesta-3,5-dien-7-one, cholestane, cholesta-3,5-diene, cholesteryl acetate and cholesteryl caprylate) were prepared and PELE simulations performed.

The starting structure for all PELE simulations was the *Aae*UPO 2.19 Å resolution crystal (PDB 2YOR) (23). As the optimum pH for peroxygenase activity is 7, the structure was accordingly prepared using Schrodinger's Protein Preparation Wizard (24) and H++ web server (25). Histidines were δ-protonated, with the exception of His-82 (ε-protonated) and His-118 and His-251 (double-protonated). All acidic residues were deprotonated, except Asp-85 that was kept in its protonated state. The eight ligands studied in this section were optimized with Jaguar (26) at the DFT/M06 level with the 6-31G** basis and a PBF implicit solvent in order to obtain their electrostatic potential atomic charges. Finally, the heme site was modeled as compound I after being fully optimized in the protein environment with quantum mechanics/molecular mechanics (QM/MM) using QSite (27).

Once the initial protein structure was prepared and ligands optimized, these were placed manually in identical positions at the entrance of the protein's binding pocket and PELE (20) simulations were performed. PELE is a Monte Carlo based algorithm that produces new configurations through a sequential ligand and protein perturbation, side chain prediction and minimization steps. New configurations are then filtered with a Metropolis acceptance test, where the energy is described with an all-atom OPLS force field (28) and a surface generalized Born solvent (29). In this way it is possible to locate and characterize local and global minima structures for the most favorable protein-ligand interactions. PELE has been successfully used in a number of ligand migration studies with both small and large substrates (30-32). In this work, PELE was set up to first drive the ligands inside the protein until the center of mass was less than 10 Å from the compound I oxygen, and after that,

simulation proceeded by allowing a free exploration of the enzyme active site by the ligand. The results presented are based on ~350 48-h trajectories for each ligand.

Sterols
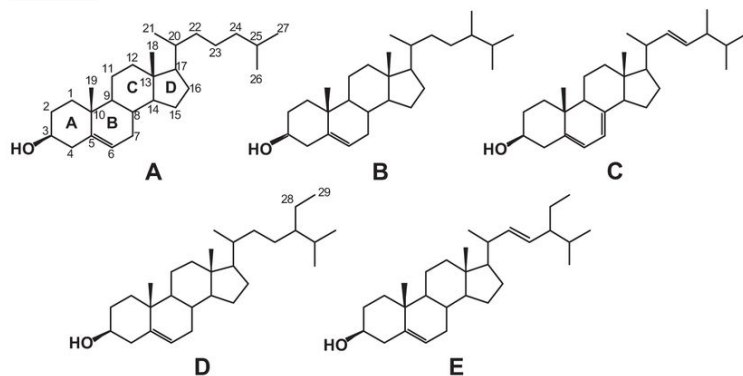
Steroid ketones
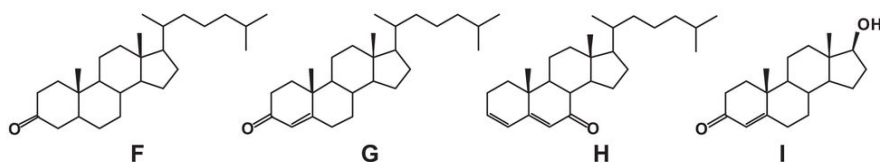
Steroid hydrocarbons
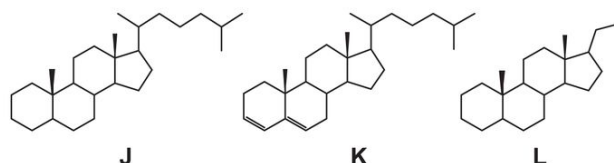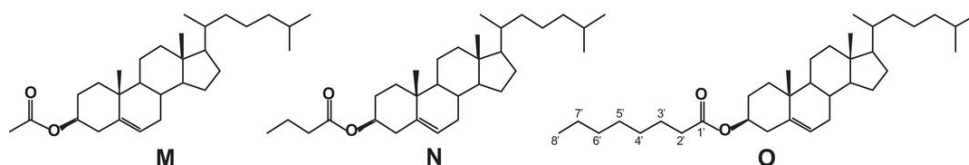

Sterol esters

FIG 1 Chemical structures of the different steroid compounds tested as substrates of the peroxygenases including: i) free sterols such as cholesterol (A), campesterol (B), ergosterol (C), sitosterol (D) and stigmasterol (E); ii) steroid ketones such as cholestan-3-one (F), 4-cholesten-3-one (G), cholesta-3,5-dien-7-one (H) and testosterone (I); iii) steroid

hydrocarbons such as cholestane (J), cholesta-3,5-diene (K) and pregnane (L); and iv) sterol esters such as cholesteryl acetate (M), cholesteryl butyrate (N) and cholesteryl caprylate (O).

RESULTS

Experimental data

The efficiency of three peroxygenases (*Aae*UPO, *Mro*UPO and r*Cci*UPO) hydroxylating fifteen steroidal compounds (including free and esterified sterols, steroid hydrocarbons and ketones) was evaluated, and the reaction products were analyzed by GC-MS.

Free sterol reactions. Free C_{27} (cholesterol), C_{28} (campesterol and ergosterol) and C_{29} (sitosterol and stigmasterol) sterols (**Fig. 1A-E**, respectively) were tested as UPO substrates (**Table 1**). Cholesterol was converted to a larger extent than the other sterols, although the individual enzymes showed differences in their performance (after 60 min): r*Cci*UPO completely (100%) transformed the substrate, followed by *Aae*UPO (64%) and *Mro*UPO (23%). The reaction products of cholesterol were oxygenated derivatives at the C24, C25 and C26 (or C27) positions of the side chain (see **Fig. 1** atom numbering), with the monohydroxylated C25 derivative predominating (Table 2). Interestingly, a strict regioselectivity was observed in the reaction of cholesterol with r*Cci*UPO, yielding 25- hydroxycholesterol as the sole reaction product.

TABLE 1 Conversion of different types of steroid compounds by the *C. cinerea* (r*Cci*UPO), *A. aegerita* (*Aae*UPO) and *M. rotula* (*Mro*UPO) peroxygenases within 60 min reaction (% of substrate transformed)

Substrate	r*Cci*UPO	*Aae*UPO	*Mro*UPO
Free sterols			
Cholesterol	100	64	23
Campesterol	30	47	10
Ergosterol	6	10	7
Sitosterol	6	13	4
Stigmasterol	2	2	5
Steroid ketones			
Cholestan-3-one	54	42	12
4-Cholesten-3-one	100	67	14
Cholesta-3,5-dien-7-one	97	57	39
Testosterone	0	0	0
Steroid hydrocarbons			
Cholestane	14	3	1
Cholesta-3,5-diene	34	18	23
Pregnane	0	0	37
Sterol esters			
Cholesteryl acetate	34	16	5
Cholesteryl butyrate	7	1	1
Cholesteryl caprylate	0	0	0

The position of the hydroxyl group was determined by the mass spectra of the TMS derivatives, as found for 25-hydroxycholesterol (**Fig. S1**). The spectrum includes the molecular ion (*m/z* 546) and a characteristic ion at *m/z* 131, resulting from C24-C25 bond cleavage, together with additional fragments. Likewise, the spectra of monohydroxylated derivatives at C24 and C26 (not shown) showed, in addition to the molecular ion at *m/z* 546, the characteristic fragment at *m/z* 131, described above, and that at *m/z* 145 and from C23- C24 bond cleavage. In the *Aae*UPO and *Mro*UPO reactions further oxidized compounds (carboxycholesterol) at C26/C27 were identified (with molecular ion at *m/z* 560, and the above characteristic fragment at *m/z* 145), though in minor amount.

TABLE 2 Abundance (relative percentage) of the different oxygenated derivatives identified by GC-MS in the reactions of free sterols with *C. cinerea* (r*Cci*UPO), *A. aegerita* (*Aae*UPO) and *M. rotula* (*Mro*UPO) peroxygenases

	24-OH	25-OH	26/27-OH	28-OH	26/27-COOH
Cholesterol					
r*Cci*UPO	-	100	-	-	-
*Aae*UPO	6	80	9	-	5
*Mro*UPO	-	89	6	-	5
Campesterol					
r*Cci*UPO	-	85	15	-	-
*Aae*UPO	-	99	1	-	-
*Mro*UPO	-	98	2	-	-
Ergosterol					
r*Cci*UPO	-	100	-	-	-
*Aae*UPO	-	100	-	-	-
*Mro*UPO	-	100	-	-	-
Sitosterol					
r*Cci*UPO	-	52	-	48	-
*Aae*UPO	-	94	-	6	-
*Mro*UPO	-	100	-	-	-
Stigmasterol					
r*Cci*UPO	-	100	-	-	-
*Aae*UPO	-	100	-	-	-
*Mro*UPO	-	100	-	-	-

The C_{28} sterols, especially ergosterol, were transformed to a lesser extent than cholesterol (**Table 1**). *Aae*UPO showed more reactivity towards campesterol and ergosterol than the other peroxygenases. The reaction of ergosterol was strictly regioselective, 25-hydroxyergosterol being the only reaction product identified in all cases (**Table 2**). The reaction with campesterol was also highly regioselective, giving the 25-hydroxyderivative as the main product, although hydroxylation at C26 was also observed (especially in the r*Cci*UPO reaction). Likewise, the three peroxygenases showed low reactivity with the C_{29} sterols, and only very low conversion rates (<13%) were attained (**Table 1**). In the reaction with stigmasterol,

only the monohydroxylated derivative at C25 was observed whereas, with sitos-terol, near the same amount of 28-hydroxysitosterol was formed by r*Cci*UPO (and a low amount by *Aae*UPO) (Table 2).

TABLE 3 Abundance (relative percentage) of the different oxygenated derivatives identified by GC-MS in the reactions of steroid ketones with *C. cinerea* (r*Cci*UPO), *A. aegerita* (*Aae*UPO) and *M. rotula* (*Mro*UPO) peroxygenases

	25-OH	26/27-OH	x-OH	x,25-diOH
Cholestan-3-one				
r*Cci*UPO	100	-	-	-
*Aae*UPO	100	-	-	-
*Mro*UPO	100	-	-	-
4-Cholesten-3-one				
r*Cci*UPO	100	-	-	-
*Aae*UPO	93	7	-	-
*Mro*UPO	100	-	-	-
Cholesta-3,5-dien-7-one				
r*Cci*UPO	93	-	-	7
*Aae*UPO	87	5	4	4
*Mro*UPO	56	-	39	5

Steroid ketone reactions. Cholestan-3-one, 4-cholesten-3-one, cholesta-3,5-dien-7-one and testosterone (**Fig. 1F-I**, respectively) were tested as UPO substrates (Table 1). r*Cci*UPO was the most efficient enzyme in transforming steroid ketones, except testosterone that was not modified by any of the three peroxygenases, followed by *Aae*UPO and *Mro*UPO. The unsaturated 4-cholesten 3-one and cholesta-3,5-dien-7-one were completely transformed by r*Cci*UPO, and the saturated cholestan-3-one was also significantly transformed (54%).

The three ketones were predominantly monohydroxylated at C25 (**Table 3**). In the case of cholestan-3-one, the regioselectivity was strict, since only the 25-monohydroxylated derivative was formed by the three enzymes, and the same was observed in the reactions of 4- cholesten-3-one with r*Cci*UPO and *Mro*UPO. However, in the reaction with *Aae*UPO, monohydroxylation at the 4-cholesten-3-one terminal positions (C26/C27) was also observed, although in minor proportion. Interestingly, in the reaction of the di-unsaturated cholesta-3,5- dien-7-one, hydroxylation in the steroidal ring was also produced being especially pronounced in the *Mro*UPO reaction.

The mass spectra of the TMS derivatives of the three 25-monohydroxylated ketones (not shown) showed the above characteristic fragment at m/z 131 (from C24-C25 bond cleavage). No molecular ions for the saturated and monounsaturated ketones were observed, although the molecular masses could be

determined from the [M-15]⁺ fragments (at *m/z* 459 and *m/z* 457, respectively). However, the mass spectrum of the 25-hydroxyderivative of the di- unsaturated ketone showed the molecular ion (*m/z* 470) in addition to the *m/z* 131 and [M-15]⁺ fragments. Likewise, the 26-hydroxyderivative of 4-cholesten-3-one, showed the molecular ion (*m/z* 472) in addition to the *m/z* 131 and [M-15]⁺ fragments. The di- unsaturated cholesta- 3,5-dien-7-one yielded the mono- and dihydroxyderivatives, with one hydroxylation at the steroidal ring. The most prominent fragment in the mass spectrum of the monohydroxylated derivative (**Fig. S2A**) corresponded to the molecular ion (at *m/z* 470). Other characteristic fragments were observed but the position of the hydroxyl group in the ring could not be determined. The mass spectrum of the dihydroxylated derivative (**Fig. S2B**) showed the characteristic fragment at *m/z* 131. Additionally, molecular ion (*m/z* 558) and other characteristic fragments were found but, like in the monohydroxylated derivative, the position of the hydroxyl group in the ring could not be determined.

Steroid hydrocarbon reactions. Cholestane, cholesta-3,5-diene and pregnane (**Fig. 1J-L**, respectively) were tested as UPO substrates (**Table 1).** Cholestane and especially, cholesta- 3,5-diene, were transformed by the three peroxygenases, r*Cci*UPO being the most efficient one. In contrast, pregnane was only modified by *Mro*UPO.

TABLE 4 Abundance (relative percentage) of the different oxygenated derivatives identified by GC-MS in the reactions of steroid hydrocarbons with *C. cinerea* (r*Cci*UPO), *A. aegerita* (*Aae*UPO) and *M. rotula* (*Mro*UPO) peroxygenases

	25-OH	x-OH	x,25-diOH	3,4-diOH-5-en	3,6-diOH-4-en	3,6,25-triOH-4-en
Cholestane						
r*Cci*UPO	100	–	–	–	–	–
*Aae*UPO	100	–	–	–	–	–
*Mro*UPO	100	–	–	–	–	–
Cholesta-3,5-diene						
r*Cci*UPO	36	–	14	3	18[a]	29
*Aae*UPO	3	–	20	35	31[b]	11
*Mro*UPO	–	–	1	36	61[c]	2
Pregnane						
r*Cci*UPO	–	–	–	–	–	–
*Aae*UPO	–	–	–	–	–	–
*Mro*UPO	–	100[d]	–	–	–	–

[a] including 3α,6α-diOH-4-en (14%) and 3β,6β-diOH-4-en (4%); [b] 3α,6α-diOH-4-en (28%) and 3β,6β-diOH-4-en (3%); [c] including 3α,6α-diOH-4-en (50%) and 3β,6β-diOH-4-en (11%); and [d] including 97% hydroxy and 3% keto derivatives (**Fig.** S3).

The reaction products of these hydrocarbons are shown in Table 4. Whereas the reaction of cholestane was 100% regioselective at C25, cholesta-3,5-diene gave several di- and trihydroxylated derivatives at the steroidal ring, in addition to the 25-hydroxyderivatives (mainly formed by r*Cci*UPO). Among the

dihydroxylated derivatives at the ring core, noteworthy are the ones with hydroxylation at C3 and C6 (**Fig. 2A and B**). In these reactions, saturation of one double bond and displacement of the other one (to C4) took place. Therefore, the 3,5-diene structure in the substrate was transformed into a 3,6-dihydroxy-4-ene structure. Likewise, the C3 and C4 dihydroxylated derivatives lost the double bond at C3 yielding a 3,4-dihydroxy-5-ene structure. The reaction of pregnane with *Mro*UPO gave several monohydroxylated derivatives at the steroid ring (97% of reaction products), together with a further oxidized derivative (3% 3-ketopregnane) (**Fig. S3**), although the position of the hydroxyl group could not be determined.
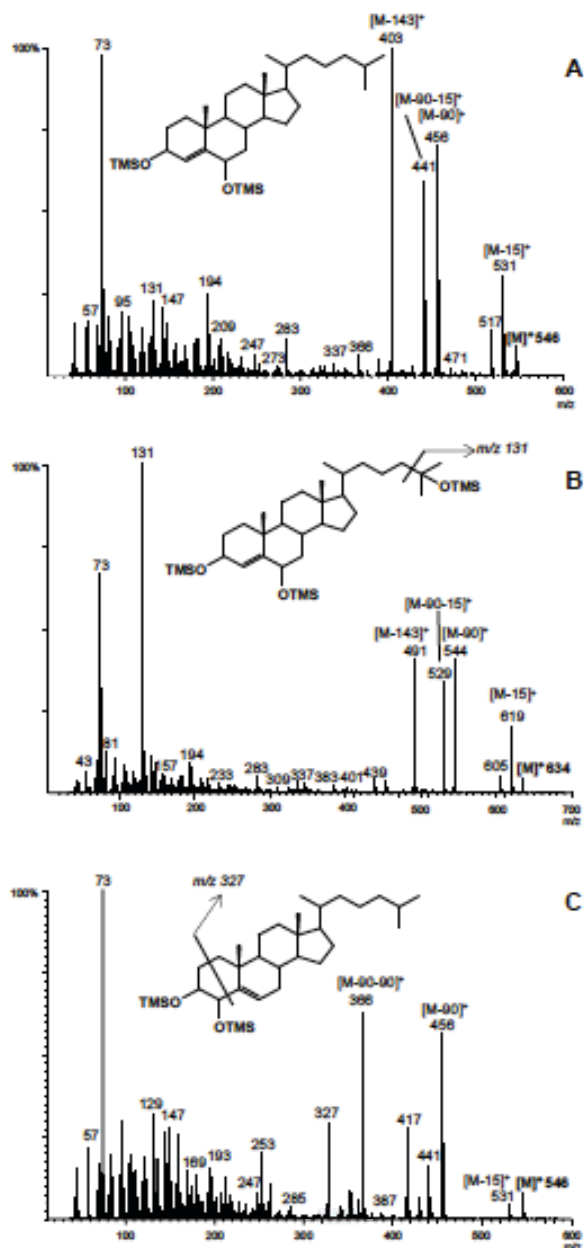
FIG 2 Mass spectra of 3,6-dihydroxycholest-4-ene (A), 3,6,25-trihydroxycholest-4-ene (B) and 3,4-dihydroxycholest-5-ene (C) from peroxygenase reactions with cholesta-3,5-diene (structure K in **Fig. 1**), as TMS derivatives.

The mass spectra of the 25-hydroxycholestane and 25-hydroxy-cholesta-3,5-diene TMS derivatives (not shown) showed the characteristic fragment at m/z 131, as well as other characteristic ([M-15]$^+$, [M-90]$^+$ and [M-90-15]$^+$) fragments. The mass spectrum of 3,6- dihydroxycholest-4-ene (**Fig. 2A**), whose formation is described above, showed molecular ion (m/z 546) and a base peak at m/z 403 considered as characteristic of the 4-ene-6-hydroxy structure (33) together with other fragments (including [M-15]$^+$, [M-90]$^+$ and [M-15-90]$^+$). The mass spectrum of 3,6,25-trihydroxycholest-4-ene (**Fig. 2B**) showed molecular ion (m/z 634) and the characteristic fragment at m/z 131, together with the three fragments mentioned above. The mass spectrum of 3,4-dihydroxycholest-4-ene (**Fig. 2C**) showed molecular ion (m/z 546) and the characteristic fragment at m/z 327, together with other fragments (three fragments mentioned above plus [M-90-90]$^+$). Finally, the mass spectrum of x,25- dihydroxycholesta-3,5-diene (not shown) showed the molecular ion (m/z 544) and the characteristic fragment at m/z 131, together with other fragments ([M-15]$^+$ and [M-90]$^+$) but the position of the hydroxyl in the steroid ring could not been determined.

Sterol ester reactions. Cholesteryl acetate, butyrate and caprylate (**Fig. 1M-O**, respectively) were tested as UPO substrates (**Table 1**). The best results were obtained with r*Cci*UPO and cholesteryl acetate reaching conversion yield of 34%. Cholesteryl caprylate was not modified by any of the three peroxygenases. In the reactions of sterol esters, the 25- hydroxyderivative was the only product identified. The mass spectra (not shown) showed the characteristic fragment at m/z 131 besides the molecular ions at m/z 544 and m/z 428, for cholesteryl butyrate and acetate, respectively.

Computational analyses

The molecular structures of eight representative steroids, and that of *Aae*UPO (PDB 2YOR) enzyme, were prepared at the optimal pH for peroxygenase reactions, and PELE simulations were performed, providing energy profiles and ligand trajectories. Since the driving criterion was to reduce the distance between the oxygen atom of the peroxygenase compound-I and the center of mass of the ligand, the fractions of simulations in which the ligand has entered either by the C25 end or the C3 end were computed. Then, the structures closest to the reactive oxygen were analyzed as described below. Further detail on energy profiles and interaction maps

for each studied steroid in the heme active site are available in supporting material (including supplemental Computational Results and **Figs. S4-S9**). A direct correlation between the percentage of substrate conversion (**Table 1**) and the percentage of entrance by C25 was observed (**Fig. 3**). In the case of cholestane, the fact that it remains 3.6 Å away from the reactive oxygen (**Fig. S7**) justifies its reduced reactivity.
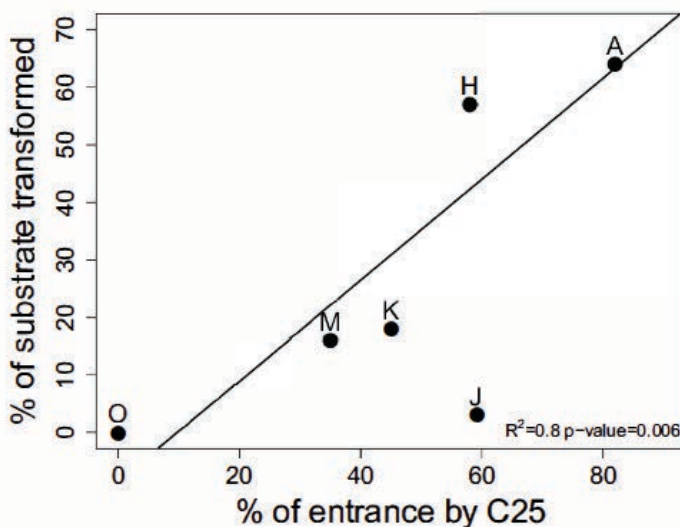

FIG 3 Correlation plot for steroid conversion rate (Table 1) and steroid (side-chain) entrance by C25 to the enzyme active site. Computational results computed from PELE diffusion simulations (20) of eight selected steroids on the molecular structure of *Aae*UPO (PDB 2YOR). For steroid (A-K) identification see Fig. 1. Cholestane (compound J) has been removed from the correlation plot (discussed in text).

A comparison of PELE simulations (binding energy *vs* distance between substrate C25 and compound-I oxygen) for side-chain hydroxylation of two free sterols and one esterified sterol (cholesterol, sitosterol and cholesteryl acetate) is shown in **Fig. 4**, which also includes the main interactions in a representative structure with the ligand at reaction distance from compound I. Cholesterol has the best fraction (82%) of entrance trajectories (**Fig. 3**). The reason for this is a very hydrophobic heme access channel of peroxygenase, which favors the entrance of the apolar C-25 side. Then, with the side chain in the active site, the molecule is further stabilized by a hydrogen bond with Glu245 on the surface of the protein (**Fig. 4A**). Reaction at C24 and C26/C27 is also possible, despite the lower reactivity of

secondary and primary carbons compared with tertiary C25, due to the favorable minima with these carbons well positioned for reaction (**Fig S4**).
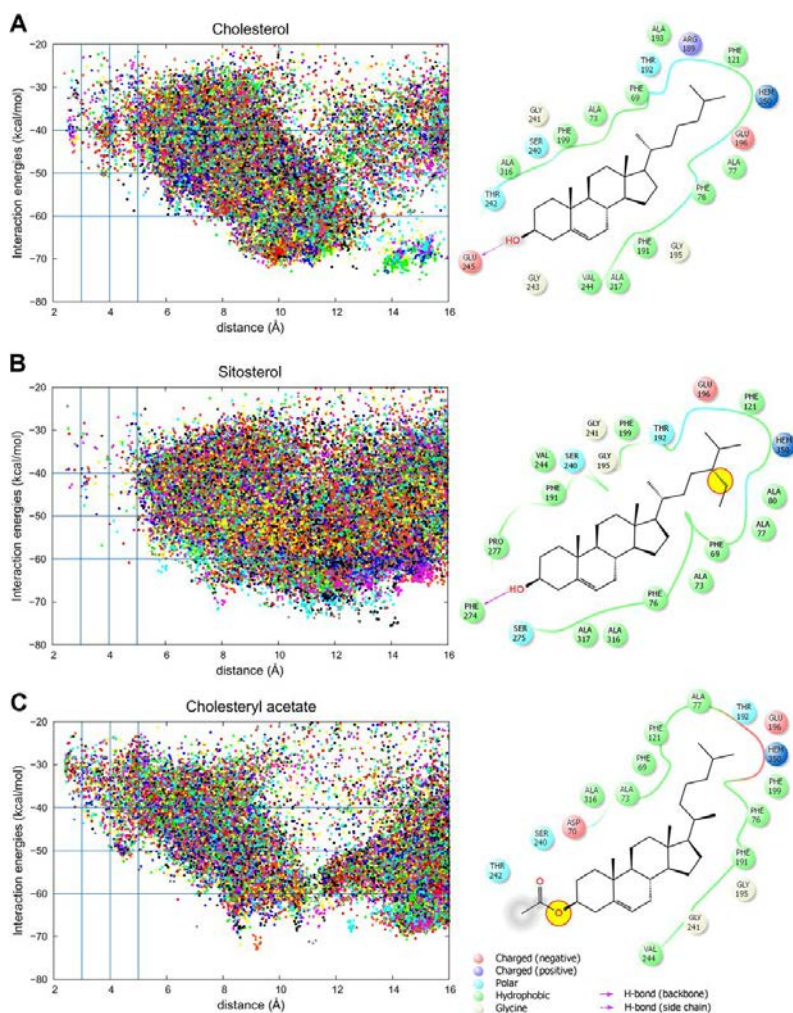

FIG 4 Results from PELE (20) simulations for cholesterol (A), sitosterol (B) and cholesteryl acetate (C) diffusion at the active site of the *Aae*UPO, including: (*left*) Plots of the energy profile *vs.* the distance between the steroid H25 and the oxygen atom in enzyme compound I; and (*right*) Main interactions between the substrate and the enzyme in a representative structure for each of the three steroids analyzed. The yellow circles identify structural differences relative to cholesterol.

Compared with cholesterol, sitosterol contains an extra ethyl group. This large substituent impedes the access of the substrate to the active site, which is reflected by the lower fraction of structures entering the protein through the C25 side (39%) compared with cholesterol (**Fig. 3**). In particular, at the active site the

C28 and C29 groups clash with Ala77 and Thr192, and reduce considerably the number of structures approaching the heme (**Fig. 4B**). In addition to C25, some reaction can be expected at C28 since the ligand also approaches to a favorable distance for reaction of this second secondary carbon (data not shown). In spite of its similar structure, the fraction of cholestan-3-one entrance trajectories (49%) is also lower than found for cholesterol (**Fig. 3**), and the distance between C25 and the compound-I oxygen is longer (**Fig. S5**). This is most probably because cholestan-3-one is unable of hydrogen bonding to Glu245. In the case of cholesta-3,5-dien-7-one, a better approach from the ring side, due to the two double bonds present (**Fig. S6**), results in the observed ring hydroxylation.

Cholesteryl acetate has an ester bond at the C3 hydroxyl and, as in the case of the above ketones, binding for C25 oxidation is no longer favored by the hydrogen bond with Glu245 (**Fig. 4C**) resulting in lower entrance trajectories (35%) than cholesterol (**Fig. 3**). Approach from the opposite side although produced did not result in ring hydroxylation. Cholesteryl caprylate has a longer acyl chain than the above cholesteryl acetate, and (as seen in **Fig. S9**) it does not approach the heme group from either side (the large hydrophobic chains would favor interactions with the protein to avoid solvent exposure).
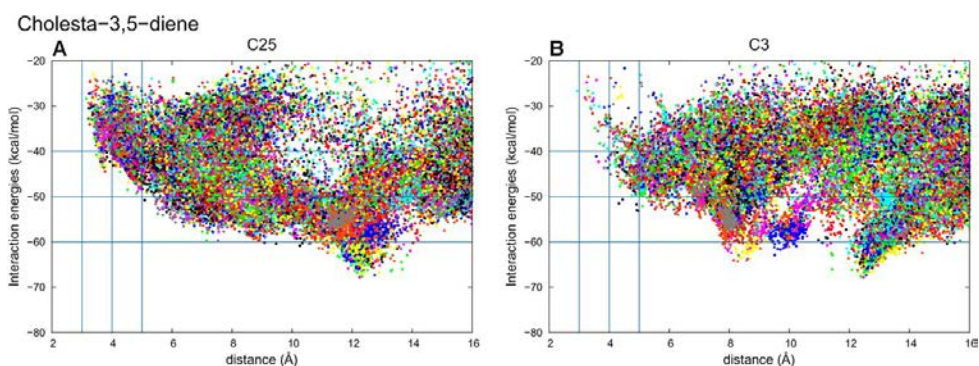

FIG 5 Results from PELE (20) simulations for cholesta-3,5-diene diffusion at the active site with C25 (A) or C3 (B) approaching the heme cofactor of the *Aae*UPO.

Finally, cholesta-3,5-diene has a very interesting reactivity, due to the presence of two double bonds at C3 and C5, and the lack of the C3 hydroxyl mentioned above, that resulted in similar entrance by both C25 and C3 positions (**Fig. 5A** and B, respectively). The latter explains the fraction of hydroxylated products at the steroid rings, many of them being additions to its double bonds. In contrast, the bulkier saturated C5 and C6 in cholestane (and consequent loss of planarity) impedes a suitable binding of the ligand when it enters by C25, and a very low number

of structures come close enough to the compound-I oxygen (**Fig. S7**), explaining the low reactivity of this compound (cholesterol is converted by 64% while only 3% for cholestane) (**Fig. 3).**

DISCUSSION

Despite significant progress in the development of efficient biocatalysts, there is still a great demand for cost-efficient and economical biotechnologies to produce valuable steroids. Most work has been dedicated to steroid modifications catalyzed by whole microbial cells (1). The present work deals with the oxyfunctionalization of steroids using fungal peroxygenases, relatively young representatives of the superfamily of heme-thiolate peroxidases, characterized by the presence of a cysteine residue as the fifth ligand of the heme iron and their peroxygenase activity (8). Several steroids differing in their structure - including free and esterified sterols, and steroid ketones and hydrocarbons - were tested as substrates of peroxygenases from three basidiomycete species. The relationships between the structure of the different substrates, the enzyme reactivity, and the reaction products obtained are discussed below. A combined experimental and computational approach shows that there is not a unique factor that determines the hydroxylation profile but instead it is a result of the sum of many structural factors, concerning both the steroid molecule and the enzyme active site. These include the polar/apolar character of C3/C7 ring groups, the presence of double bonds in the steroidal rings, and the length and character of the alkyl C17 chain, which affect the ligand entrance ratio by the C25 side.

Relationships between steroid structure and conversion yield

Alkyl chain at C17. The influence of the steroid alkyl side-chain substitution and unsaturation was shown by the different transformation extents obtained for C_{27}, C_{28} and C_{29} sterols. The C_{27} cholesterol was completely transformed by r*Cci*UPO, while the other sterols were transformed to a lesser extent by this and the two other peroxygenases. The presence of a methyl or ethyl group at C24 makes it progressively more difficult to hydroxylate campesterol and sitosterol as illustrated in **Fig. 6**. Moreover, the double bound at C22, in addition to the methyl or ethyl group at C24, made ergosterol and stigmasterol more difficult to be hydroxylated (by the three peroxygenases) compared with the above campesterol and sitosterol. The influence of a shorter side-chain was shown by comparing hydroxylation of pregnane (with an ethyl group at C17) and other steroid hydrocarbons (e.g. cholestane). The presence of a two-carbon (instead of eight-carbon) alkyl chain prevents

hydroxylation by r*Cci*UPO and *Aae*UPO and, in contrast, promotes hydroxylation by *Mro*UPO, which takes place in the steroidal ring (not in the side-chain). Finally, the absence of an alkyl chain at C17 like in testosterone (having instead a hydroxyl group) completely prevents its hydroxylation by the three fungal peroxygenases (compared with up to 100% conversion of 4-cholesten-3- one)
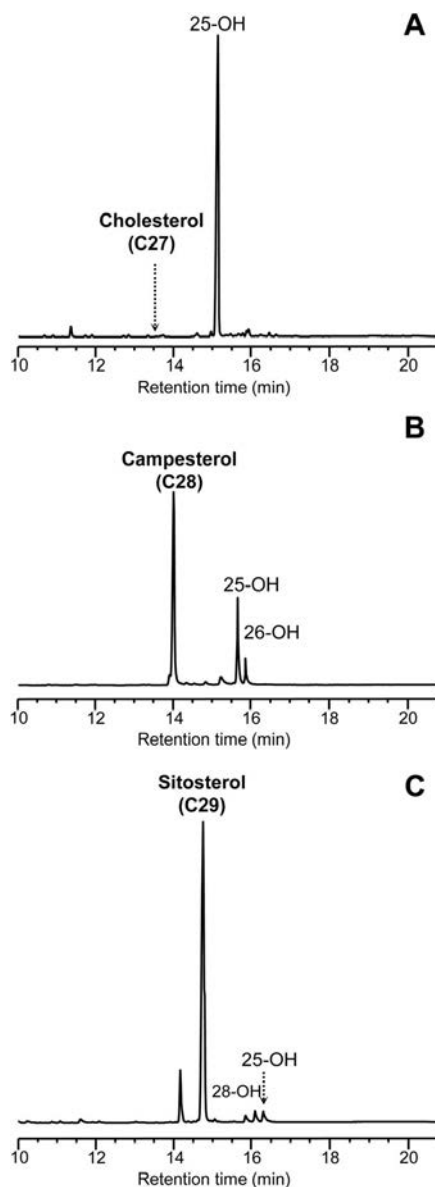

FIG 6 GC-MS analysis of the r*Cci*UPO reaction (at 60 min) with cholesterol (A), campesterol (B) and sitosterol (C) (see **Fig. 1**) showing the remaining substrates and the monohydroxylated derivatives at C25, C26 and C28 of the side chain.

The good conversion produced in the reactions of peroxygenases with cholesterol and the differences observed in the conversion of sterols with different alkyl chain size, like sitosterol, are explained by computational results. In general terms, the polarity of the substituent group at the C3 has a positive correlation with the ligand rate entrance by C25 side. This is due to the hydrophobic character of the protein access channel to the heme that contains numerous phenylalanine residues. Moreover, the presence of a hydroxyl group in C3 (of cholesterol and other sterols), increases the ligand rate entrance by C25 side and stabilizes the ligand in a suitable catalytic position (a good catalytic position is assumed when the distance between the hydrogen in C25 and the compound I oxygen is around 2.4 Å) due to a hydrogen bond with Glu245. These two factors correlate perfectly with the high activity shown for cholesterol. However, although sitosterol has also a hydroxyl group in C3, it is not able to access the heme site by the C25 end in the same way as cholesterol due to the bulkier alkyl chain in C17 that clashes with residues in the heme access channel (namely Thr192, Ala77, Phe69, Phe121 and Phe199).

Oxidation/unsaturation degree of the rings. Steroids differing in the oxidation state of the ring were tested including steroid hydrocarbons, sterols and steroid ketones. Whereas r*Cci*UPO completely converted both cholesterol and 4-cholesten-3-one, *Aae*UPO and *Mro*UPO oxidized the ketone with lower efficiency. The contrary was reported for P450s (e.g. CYP27A1), which hydroxylated steroids with a 3-oxo-Δ^4 structure at a much higher rate than those with a 3β-hydroxy-Δ^5 structure (34). Therefore, the structural differences between 4- cholesten-3-one and cholesterol – that are the planarity of the sterol A-ring and the electronic properties resulting from a keto *versus* a hydroxyl substituent at C3 – seemingly affect the peroxygenase activity. The computational results revealed that the polar character of the hydroxyl and oxo groups in the sterols and steroid ketones affects in a positive way the ligand entrance rate by C25, which correlates with the high activity that these compounds show compared with the corresponding steroid hydrocarbons.

On the other hand, although differences were observed between the three peroxygenases, the presence and number of double bonds in the rings of steroid hydrocarbons and ketones significantly increases their hydroxylation degree. This is illustrated in **Fig. 7**, where better *Aae*UPO conversion of cholesta-3,5-diene and cholesta-3,5-dien-7-one, compared with the corresponding saturated molecules (cholestane and cholestan-3-one), is shown. The C25 side entrance and the placement of the steroid hydrocarbons in a proper catalytic position are determined by the presence or absence of double bonds. These reduce the volume of the steroid ring facilitating its entrance in the active site and provide proper curvature to the substrate for an optimal approach to the heme. Therefore, both the apolar nature of

the two extremes of the molecule and the presence of double bonds in cholesta-3,5-diene facilitate the access to the heme from both sides, C3 and C25, and its correct placement for reaction. However, neither the substrate entrance nor its placement in the active site are as suitable as in sterols and steroid ketones. In the case of cholestane the entrance to the heme site is hampered by the larger sp3 groups at C3, C4, C6 and C7. This can be better observed by comparing the energy profiles of this compound (**Fig. S7**) to cholesta-3,5-diene (**Fig. S8**) and observing the very different occupation of the binding site. In addition, the higher reactivity of alkenes as compared to alkanes lead to addition of hydroxyl groups to the double bonds in the steroidal rings in cholesta-3,5-diene which is not observed in cholestane. Concerning the two ketones compared in **Fig. 7C and D**, although both have a similar ratio for C25 entrance, cholestan-3- one can only be hydroxylated at the C25, while cholesta-3,5-dien-7-one also shows some additional hydroxylation due to the higher reactivity of alkenes mentioned above. Moreover, as in the case of the steroid hydrocarbons, the two double bonds in the steroid ketone ring will facilitate the entrance and proper positioning of the substrate at the peroxygenase active site.

Due to the above effects, although the conversion rates of cholestan-3-one are lower than those of cholesterol (the best steroid substrate of the basidiomycete peroxygenases) they are increased by the presence of double bonds in the steroid ring, and cholestan-3,5-dien-7-one hydroxylation rates are in the same order of those obtained for cholesterol.

Esterified hydroxyl group at C3. With the aim of investigating whether the hydroxyl group at C3, in free form, is necessary for maximal peroxygenase activity, three esters of cholesterol with organic acids of different chain length were assayed as substrates. The experimental results revealed that the presence of an esterifying group hampers substrate conversion, and that this effect is directly correlated with increasing chain length of the fatty acid. In sterol esters, only hydroxylation (at C25) of the side-chain of the sterol was observed and not at the acyl moiety, in spite that *Aae*UPO and r*Cci*UPO were found to efficiently hydroxylate fatty acids (11, 18).

The computational analyses revealed that in the sterol esters the C25 entrance rate is decreased with respect to cholesterol due to the presence of hydrophobic alkyl chains at the ester site. In the case of cholesteryl caprylate, which is not transformed by any of the peroxygenases, the long unsubstituted hydrocarbon chain tends to "anchor" the substrate away from the heme site by interacting with different groups on the protein surface. Since both ends are hydrophobic,

interactions with the protein minimize the solvent exposure but restrain the ligand from reaching the heme, which explains the lack of activity for this compound.
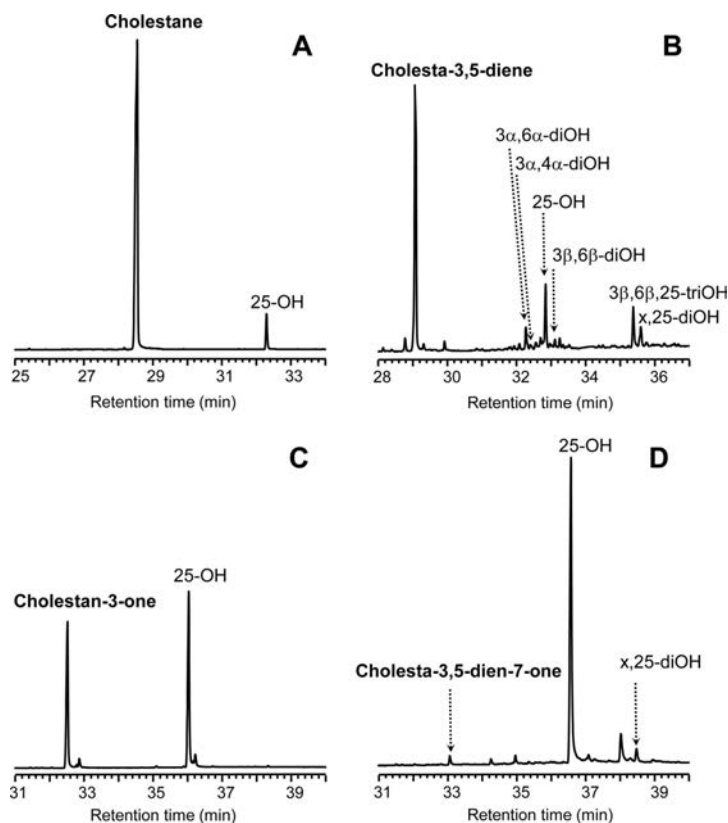

FIG 7 GC-MS analysis of the r*Cci*UPO reaction (at 60 min) with cholestane (A), cholesta-3,5-diene (B), cholestan-3-one (C) and cholesta-3,5-diene-7-one (D) (see Fig. 1) showing the remaining substrates and the mono-, di- and trihydroxylated derivatives included in Table 4 (x denotes the unknown position of hydroxylation in the ring).

Relationships between steroid structure and regioselectivity

In general, the reactions of the three fungal peroxygenases with the different steroids proceed with remarkable regioselectivity, and result in the formation of the 25-hydroxyderivatives as the main products. Noteworthy is the complete conversion of cholesterol by r*Cci*UPO to produce 25-hydroxycholesterol, since this compound is widely used, displaying an array of pharmacological actions *in vitro* and in cell-based systems (35) as discussed in next section. Different alkyl chains at C17 do seemingly not influence the regioselectivity in *Aae*UPO and *Mro*UPO reactions (see below) but affect that of r*Cci*UPO, since cholesterol is

C25- hydroxylated with a strict (100%) regioselectivity whereas, for campesterol and sitosterol, hydroxylation at terminal C26/C27 (15%) and subterminal C28 (48%), respectively, was also observed. On the other hand, the double bound at C22 in ergosterol and stigmasterol seemingly influences the strict regioselectivity of the hydroxylation at C25. Minor hydroxylation at the terminal positions of the steroid branched side-chain by *Aae*UPO and *Mro*UPO also leads in some cases (e.g. cholesterol) to formation of the C26/C27-carboxylated derivatives, in addition to the C26/C27-hydroxylated ones. This is because of the further oxidation (successive hydroxylation and dehydration reactions) of the monohydroxylated products, as previously demonstrated by the authors in ^{18}O-labeling oxygenation studies with *Aae*UPO and $H_2^{18}O_2$ (11).

Hydroxylation at positions other than C25 occurs in cases that show a high (energetically favorable) occupation of binding positions near any of these centers, as demonstrated for *Aae*UPO by computational studies. In particular, the case of cholesterol hydroxylation at C24 is due to interactions between the substrate and the protein that include a local minima for C24, which is slightly more favorable than for C25. Hydroxylation at C25 (tertiary C-H bond) is electronically preferred by its lower C-H bond strength compared with that of the secondary and primary C-H bonds (36). In contrast, steric constraints rather than the differential electronic properties of the C-H bonds have been reported to govern the regioselectivity of the reaction of steroids (cholesterol and 4-cholesten-3-one) with P450s (CYP125A1) where terminal oxidation was preferred (37).

Interestingly, in addition to hydroxylation at the side-chain, some hydroxylated derivatives at the steroid core are formed when oxygen groups are lacking at C3, such as in cholesta-3,5- dione, cholesta-3,5-diene or pregnane. Indeed, the presence of a hydroxyl or oxo group at this position seems to prevent the hydroxylation of the steroidal core. On the other hand, the presence of two conjugated double bonds in the ring seems to favor the hydroxylation at the steroidal core, especially by *Aae*UPO and *Mro*UPO, as evidenced in the reaction of cholesta- 3,5-diene yielding 3,6-dihydroxycholest-4-ene and 3,4-dihydroxycholest-5-ene mentioned above. The mechanism of this double (at C3 and C6 or at C3 and C4) hydroxylation would imply the displacement of one of the cholesta-3,5-diene double bonds to the intermediate (C4 or C5) position and the saturation of the other.

Significance of side-chain versus ring hydroxylation

It has been shown that the reactions of the peroxygenases from the three different basidiomycetes occurred generally with noteworthy regioselectivity, and

hydroxylation at the side chain over the steroidal ring was preferred, with the 25-hydroxyderivatives predominating. From a biochemical point of view, side-chain and steroid nucleus oxidation have been confirmed to be independent processes, at least in some microorganisms (1). Side- chain hydroxylation is usually a pre-step in side-chain degradation in most microbial systems that oxidize steroids.

The initial step of the side-chain oxidation of sterols (and other C_{27} steroids) is hydroxylation at C26 or C27, followed by a complex sequence of reactions involving different enzymes (some of them still unidentified) that produce elimination of side chain at C17 and then steroidal nucleus oxidation. However, further oxidative degradation on the C17 side chain may not occur when hydroxylation is produced at C25 (instead of C26 or C27). This has been observed using 25-hydroxylated cholesterol as substrate (38) concluding that C25 hydroxylation must be a distinct reaction from the usual cholesterol side-chain degradation pathway. Indeed, the authors do not consider that steroid metabolism is a specific function of UPOs, not as much as it is the oxidation of alkanes, alkenes, ethers or aromatics. UPOs rather represent a catalytic system ("extracellular liver") that, irrespective of a particular structure, can non-specifically oxidize diverse compounds emerging in the fungus microenvironment. In this context, unspecific detoxification is surely more important than specific degradation.

Interestingly, a previously unrecognized biological (antiviral) role for 25-hydroxycholesterol, which is selectively formed by the r*Cci*UPO as discussed above, has been recently highlighted (39). Indeed, the above study revealed 25-hydroxycholesterol as the only oxysterol notably upregulated following macrophage infection or activation by interferons. This antiviral effect is specific to 25-hydroxycholesterol because other related (enzymatically generated) oxysterols, such as 19-hydroxychoesterol and 7α-hydroxycholesterol, fail to repress viral infection. 25-hydroxycholesterol has been reported to have a high potency to inhibit a broad spectrum of viruses from high to low physiological concentrations depending on lipid conditions and virus-host cell system. Likewise, other recent studies also revealed 25- hydroxycholesterol as potential antiviral therapeutic (40). Independent of its known regulatory effect on metabolism, 25-hydroxycholesterol impairs viral entry at the virus-cell fusion step by inducing cellular membrane changes. Because 25-hydroxycholesterol can permeate through membranes, it likely modifies cellular membranes to exert its antiviral effect. Moreover, in addition to interfering with viral entry and replication, 25- hydroxycholesterol also amplifies the activation of immune cells and increases the production of immune mediators (41).

Final remarks

The first crystal structure of a fungal peroxygenase (23), from the basidiomycete *A. aegerita* (*Aae*UPO), was used in the computational simulations included in the present study. The availability of more peroxygenase crystal structures, including those from *C. cinerea* and *M. rotula*, will provide the opportunity to correlate the different regioselectivities in steroid oxygenation described here with the architecture of the active site, and other structural features. Such information will permit us to engineer these self-sufficient monooxygenases (whose activation just depends on a peroxide source) as new and robust industrial biocatalysts for the pharmaceutical and fine chemical sectors.

In conclusion, in the present work, the ability of three different fungal peroxygenases to catalyze the regioselective hydroxylation of a variety of steroids is shown for the first time. Generally, hydroxylation at the side chain over the steroid ring was preferred, 25-hydroxyderivatives being the main products formed. It has been observed that some structural features of the steroid substrates, such as presence and length of the alkyl chain at C17 and the presence of oxygen groups and/or conjugated double bonds in the steroid ring system, influence substrate conversion and guide regioselectivity of the reactions. A better understanding of the mechanistic aspects of hydroxylation in combination with the development of suitable biocatalysts would be the basis for efficient steroid hydroxylation bioprocesses using new enzymes with self-sufficient mono(per)oxygenase activities.

ACKNOWLEDGEMENTS

REFERENCES

1. Donova MV and Egorova OV. 2012. Microbial steroid transformations: current state and prospects. *Appl.Microbiol.Biotechnol*. 94:1423-1447.

2. *Beneventi E, Ottolina G, Carrea G, Panzeri W, Fronza G, and Lau PCK. 2009. Enzymatic Baeyer-Villiger oxidation of steroids with cyclopentadecanone monooxygenase.* J.Mol.Catal.B-Enzym. *58:164-168.*

3. Ortiz de Montellano, PR. 2005. Cytochrome P450: Structure, mechanism, and biochemistry. Kluwer Academic/Plenum, New York.

4.   Urlacher VB and Girhard M. 2012. Cytochrome P450 monooxygenases: an up-
      date on perspectives for synthetic application. *Trends Biotechnol*. 30:26-36.

5.   van Beilen JB and Funhoff EG. 2005. Expanding the alkane oxygenase
      toolbox: new enzymes and applications. *Curr.Opin.Biotechnol.* 16:308-314.

6.   Li, H. 2001. Cytochrome P450, p. 267-282. *In*: A. Messerschmidt, R. Huber,
      T. L. Poulos, and K. Wieghardt (eds.), Handbook of metalloproteins., vol. 1.
      Wiley, Baffins Lane, UK.

7.   Ullrich R, Nuske J, Scheibner K, Spantzel J, and Hofrichter M. 2004. Novel
      haloperoxidase from the agaric basidiomycete *Agrocybe aegerita* oxidizes
      aryl alcohols and aldehydes. *Appl.Environ.Microbiol.* 70:4575-4581.

8.   Hofrichter M and Ullrich R. 2014. Oxidations catalyzed by fungal peroxygen-
      ases. *Curr.Opin.Chem.Biol.* 19:116-125.

9.   Bernhardt R. 2006. Cytochromes P450 as versatile biocatalysts. *J.Biotechnol*.
      124:128-145.

10.   Hofrichter M, Ullrich R, Pecyna MJ, Liers C, and Lundell T. 2010. New and
      classic families of secreted fungal heme peroxidases. *Appl.Microbiol.Biotech-
      nol*. 87:871-897.

11.   *Gutiérrez A, Babot ED, Ullrich R, Hofrichter M, Martínez AT, and del Río JC.
      2011. Regioselective oxygenation of fatty acids, fatty alcohols and other ali-
      phatic compounds by a basidiomycete heme-thiolate peroxidase.* Arch.Bio-
      chem.Biophys. *514:33-43.*

12.   *Peter S, Kinne M, Wang X, Ullrich R, Kayser G, Groves JT, and Hofrichter
      M. 2011. Selective hydroxylation of alkanes by an extracellular fungal perox-
      ygenase.* FEBS J. *278:3667-3675.*

13.   Anh DH, Ullrich R, Benndorf D, Svatos A, Muck A, and Hofrichter M. 2007.
      The coprophilous mushroom *Coprinus radians* secretes a haloperoxidase that
      catalyzes aromatic peroxygenation. *Appl.Environ.Microbiol.* 73:5477-5485.

14.   Gröbe G, Ullrich M, Pecyna M, Kapturska D, Friedrich S, Hofrichter M, and
      Scheibner K. 2011. High-yield production of aromatic peroxygenase by the ag-
      aric fungus *Marasmius rotula* . *AMB Express.* 1:31-42.

15.   Pecyna MJ, Ullrich R, Bittner B, Clemens A, Scheibner K, Schubert R, and
      Hofrichter M. 2009. Molecular characterization of aromatic peroxygenase
      from *Agrocybe aegerita*. *Appl.Microbiol.Biotechnol*. 84:885-897.

16.   Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martínez
      AT, Otillar R, Spatafora JW, Yadav JS, Aerts A, Benoit I, Boyd A, Carlson

A, Copeland A, Coutinho PM, de Vries RP, Ferreira P, Findley K, Foster B, Gaskell J, Glotzer D, Górecki P, Heitman J, Hesse C, Hori C, Igarashi K, Jurgens JA, Kallen N, Kersten P, Kohler A, Kües U, Kumar TKA, Kuo A, LaButti K, Larrondo LF, Lindquist E, Ling A, Lombard V, Lucas S, Lundell T, Martin R, McLaughlin DJ, Morgenstern I, Morin E, Murat C, Nolan M, Ohm RA, Patyshakuliyeva A, Rokas A, Ruiz-Dueñas FJ, Sabat G, Salamov A, Samejima M, Schmutz J, Slot JC, St.John F, Stenlid J, Sun H, Sun S, Syed K, Tsang A, Wiebenga A, Young D, Pisabarro A, Eastwood DC, Martin F, Cullen D, Grigoriev IV, and Hibbett DS. 2012. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science.* 336:1715-1719.

17. Stajich JE, Wilke SK, Ahren D, Au CH, Birren BW, Borodovsky M, Burns C, Canback B, Casselton LA, Cheng CK, Deng JX, Dietrich FS, Fargo DC, Farman ML, Gathman AC, Goldberg J, Guigo R, Hoegger PJ, Hooker JB, Huggins A, James TY, Kamada T, Kilaru S, Kodira C, Kues U, Kupfert D, Kwan HS, Lomsadze A, Li WX, Lilly WW, Ma LJ, Mackey AJ, Manning G, Martin F, Muraguchi H, Natvig DO, Palmerini H, Ramesh MA, Rehmeyer CJ, Roe BA, Shenoy N, Stanke M, Ter Hovhannisyan V, Tunlid A, Velagapudi R, Vision TJ, Zeng QD, Zolan ME, and Pukkila PJ. 2010. Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proc.Natl.Acad.Sci.USA.* 107:11889-11894.

18. *Babot ED, del Río JC, Kalum L, Martínez AT, and Gutiérrez A. 2013. Oxyfunctionalization of aliphatic compounds by a recombinant peroxygenase from Coprinopsis cinerea .* Biotechnol.Bioeng. *110:2332.*

19. Cossins BP, Hosseini A, and Guallar V. 2012. Exploration of protein conformational change with PELE and meta-dynamics. *J.Chem.Theory Comput.* 8:959-965.

20. Borrelli KW, Vitalis A, Alcantara R, and Guallar V. 2005. PELE: Protein energy landscape exploration. A novel Monte Carlo based technique. *J.Chem.Theory Comput.* 1:1304-1311.

21. Ullrich R, Liers C, Schimpke S, and Hofrichter M. 2009. Purification of homogeneous forms of fungal peroxygenase. *Biotechnol.J.* 4:1619-1626.

22. Gutiérrez A, del Río JC, González-Vila FJ, and Martín F. 1998. Analysis of lipophilic extractives from wood and pitch deposits by solid-phase extraction and gas chromatography. *J.Chromatogr.* 823:449-455.

23. Piontek K, Strittmatter E, Ullrich R, Grobe G, Pecyna MJ, Kluge M, Scheibner

K, Hofrichter M, and Plattner DA. 2013. Structural basis of substrate conversion in a new aromatic peroxygenase: Cytochrome P450 functionality with benefits. *J.Biol.Chem.* 288:34767-34776.

24. *Sastry GM, Adzhigirey M, Day T, Annabhimoju R, and Sherman W. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments.* Journal of Computer-Aided Molecular Design. *27:221-234.*

25. Anandakrishnan R, Aguilar B, and Onufriev AV. 2012. H++3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Res. 40:W537-W541.

26. Schrödinger. 2011. Jaguar 7.8. LCC, New York.

27. Schrödinger. 2011. QSite 5.7. LCC, New York.

28. Kaminski GA, Friesner RA, Tirado-Rives J, and Jorgensen WL. 2001. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J.Phys.Chem.B.* 105:6474-6487.

29. Bashford D and Case DA. 2000. Generalized born models of macromolecular solvation effects. *Annual Review of Physical Chemistry.* 51:129-152.

30. Linde D, Pogni R, Cañellas M, Lucas F, Guallar V, Baratto MC, Sinicropi A, Sáez-Jiménez V, Coscolín C, Romero A, Medrano FJ, Ruiz-Dueñas FJ, and Martínez AT. 2014. Catalytic surface radical in dye-decolorizing peroxidase: A computational, spectroscopic and directed mutagenesis study. *Biochem.J.* on-line doi:10.1042/BJ20141211.

31. Hernández-Ortega A, Lucas F, Ferreira P, Medina M, Guallar V, and Martínez AT. 2011. Modulating O2 reactivity in a fungal flavoenzyme: Involvement of aryl- alcohol oxidase Phe-501 contiguous to catalytic histidine. *J.Biol.Chem.* 286:41105- 41114.

32. Lucas MF and Guallar V. 2012. An Atomistic View on Human Hemoglobin Carbon Monoxide Migration Processes. *Biophys.J.* 102:887-896.

33. Brooks CJW, Henderso W, and Steel G. 1973. Use of trimethylsilyl ethers in characterization of natural sterols and steroid diols by gas chromatography mass spectrometry. *Biochim.Biophys.Acta.* 296:431-445.

34. Norlin M, von Bahr S, Bjorkhem I, and Wikvall K. 2003. On the substrate specificity of human CYP27A1: implications for bile acid and cholestanol formation. *J.Lipid Res.* 44:1515-1522.

35. Diczfalusy U. 2013. On the formation and possible biological role of 25- hydroxycholesterol. *Biochimie.* 95:455-460.

36. *Johnston JB, Ouellet H, Podust LM, and Ortiz de Montellano PR. 2011. Structural control of cytochrome P450-catalyzed -hydroxylation.* Arch.Biochem.Biophys. *507:86-94.*

37. Ouellet H, Guan SH, Johnston JB, Chow ED, Kells PM, Burlingame AL, Cox JS, Podust LM, and de Montellano PR. 2010. *Mycobacterium tuberculosis* CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3- one. *Mol.Microbiol.* 77:730-742.

38. Wang KC, Wang PH, and Lee SS. 1997. Microbial transformation of protopanaxadiol and protopanaxatriol derivatives with *Mycobacterium* sp. (NRRL B- 3805). *J.Nat.Prod.* 60:1236-1241.

39. Blanc M, Hsieh WY, Robertson KA, Kropp KA, Forster T, Shui GH, Lacaze P, Watterson S, Griffiths SJ, Spann NJ, Meljon A, Talbot S, Krishnan K, Covey DF, Wenk MR, Craigon M, Ruzsics Z, Haas J, Angulo A, Griffiths WJ, Glass CK, Wang YQ, and Ghazal P. 2013. The Transcription Factor STAT-1 Couples Macrophage Synthesis of 25-Hydroxycholesterol to the Interferon Antiviral Response. *Immunity.* 38:106-118.

40. Liu SY, Aliyari R, Chikere K, Li GM, Marsden MD, Smith JK, Pernet O, Guo HT, Nusbaum R, Zack JA, Freiberg AN, Su LS, Lee B, and Cheng GH. 2013. Interferon-Inducible Cholesterol-25-Hydroxylase Broadly Inhibits Viral Entry by Production of 25-Hydroxycholesterol. *Immunity.* 38:92-105.

41. Gold ES, Diercks AH, Podolsky I, Podyminogin RL, Askovich PS, Treuting PM, and Aderem A. 2014. 25-Hydroxycholesterol acts as an amplifier of inflammatory signaling. *Proc.Natl.Acad.Sci.USA.* 111:10666-10671.

Supplemental Material

Steroid hydroxylation by basidiomycete peroxygenases: A combined experimental and computational study

Esteban D. Babot,[a] José C. del Río,[a] Marina Cañellas,[b,c] Ferran Sancho,[b,c] Fátima Lucas,[b] Víctor Guallar,[b,d] L. Kalum,[e] Henrik Lund,[e] Glenn Gröbe,[f] Katrin Scheibner,[f] René Ullrich,[g] Martin Hofrichter,[g] Angel T. Martínez[h] and Ana Gutiérrez[a*]

[a]Instituto de Recursos Naturales y Agrobiología de Sevilla, CSIC, Reina Mercedes 10, E-41012 Seville, Spain

[b]Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain,

[c]Anaxomics Biotech, Balmes 89, E-08008 Barcelona, Spain

[d]ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain

[e]Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark

[f]JenaBios GmbH, Orlaweg 2, 00743 Jena, Germany

[g]TU Dresden, Department of Bio- and Environmental Sciences, Markt 23, 02763 Zittau, Germany

[h]Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain

The Supplemental Material include Supplemental Computational Results, and Supplemental Figures showing mass spectrum of 25-hydroxycholesterol (**Fig. S1**), mass spectra of x-hydroxycholesta-3,5-dien-7-one and 25,x-dihydroxycholesta-3,5-dien-7-one (**Fig. S2**), and chromatogram of the pregnane reaction (**Fig. S3**); together with PELE simulations for cholesterol (**Fig. S4**), cholestan-3-one (**Fig. S5**), cholesta-3,5-dien-7-one (**Fig. S6**), cholestane (**Fig. S7**), cholesta-3,4-diene (**Fig. S8**) and cholesteryl caprylate (**Fig. S9**).

Supplemental Computational Results

Free and esterified sterols

Cholesterol (compound A in **Fig. 1**) has the best fraction of trajectories entering by the C25 side (82%) (**Fig. 3**). The reason for this is a hydrophobic entrance to the heme side which favors the entrance by the apolar side of the ligand. Then in the active site the molecule is further stabilized by a hydrogen bond with Glu245 on the surface of the protein (**Fig. 4A**). Reactions at C24 (**Fig. S4A**) and C26/C27 (**Fig. S4C and D**) are possible despite the less favorable breaking of secondary and primary C-H bonds due to the presence of minima with these hydrogens well positioned for reaction with compound I. In particular C24 approaches the heme with a more favorable interaction energy than C25.

Sitosterol (compound D in **Fig. 1**) contains an extra ethyl group at C24. This large substituent impedes the entrance of the substrate in the active site, which is reflected by the lower fraction of structures entering the protein through C25 side (39% opposed to 82% for compound A) (**Fig. 3**). In particular, at the active site the C28 and C29 groups clash with Ala77 and Thr192 which reduces considerably the number of structures capable of approaching the heme at a reactive distance (**Fig. 4B**). Some reaction can nevertheless be expected at C28 due to the fact that simulations show that the ligand approaches the heme to a favorable distance for reaction at this secondary carbon.

---

* Address correspondence to Ana Gutiérrez, anagu@irnase.csic.es

Steroid ketones

Cholestan-3-one (compound F in **Fig. 1**) has a ketone in C3 instead of an alcohol and a single bond between C5 and C6. This ligand is not able to enter the binding pocket as well as compound A (**Fig. S5**). First the fraction of entrances by the C25 side is reduced (from 82% to 49%) and second the distance between the reactive hydrogen and the oxygen atom in cholestan-3-one is slightly longer (2.5 Å in A to 2.8 Å in F). This will have an effect in the reactivity of the compound and in particular will favor reaction at the most labile hydrogen (H25 being tertiary) since the number of reactive positions is considerably reduced (when compared to cholesterol). Interaction of the ligand in the active site is less favorable than cholesterol due to the fact that cholestan-3-one is incapable of hydrogen bonding to Glu245.

Cholesta-3,5-dien-7-one (compound H in **Fig. 1**) does not have the alcohol in C3, instead it has a ketone in C7 and an extra double bond at C3-C4. This molecule has an identical activity as cholesterol (57% substrate conversion vs. 64). Here the entrance by the C25 side is reduced (from 82% in cholesterol to 58% in cholesta-3,5-dien-7-one) probably due to the apolar nature of the steroid ring on the external side (C3) (**Fig. S6**). This also justifies reaction observed in the ring due to the presence of the double bond and C3 which can place itself correctly in the active site. However the hydrogen present on a carbon atom with an sp^2 hybridization has lower reactivity (similar to a primary carbon) which can explain the low percentage of product at the steroid ring.

Steroid hydrocarbons

Cholestane (compound J in **Fig. 1**) lacks the polar alcohol group in C3, which increases the fraction of entrance by this side of the ligand, but most importantly the bulkier saturated C5 and C6 (and consequent loss of planarity) impedes a suitable placement of the ligand in the binding site when it enters by C25. For this reason a low number of structures come close enough to the oxygen atom in cholestane (**Fig. S7**) which explains the low reactivity of this compound (cholesterol is converted by 64% while only 3% for cholestane).

Cholesta-3,5-diene (compound K in **Fig. 1**) has a very interesting reactivity due to the presence of two double bonds at C3 and C5. Also the lack of any polar group leads to a less favorable entrance by C25 (82% for A and 45% for K) (**Fig. 3**). This compound should also be compared to cholesta-3,5-dien-7-one as they only differ by the presence of a ketone group at C7 (**Figs. S6** and **S8**). This group has a considerable effect in reactivity. First it decreases the fraction of entrances by C3 (55% for K vs. 42% for H), thus increasing the conversion of the reactants to products (18% for K 57% and for H). Second, the more favorable entrance by C3 for cholesta-3,5-diene increases the fraction of hydroxylated products at the steroid rings. It is also interesting to note that many of these products are additions to the double bond.

Sterol esters

Cholesteryl acetate (compound M in **Fig. 1**) has almost the same structure as cholesterol but with an ester instead of a hydroxyl group in the C3. For this reason, the entrance to the protein by this side (**Fig. 3**) is favoured compared to cholesterol, while the opposite occurs at the C25 side (82% for cholesterol and 35% for cholesteryl acetate) (**Fig. 3**). Moreover, the binding position of the ligand in the C25 oxidation is no longer favoured by the hydrogen bond with Glu245 (**Fig. 4**).

Cholesteryl caprylate (compound O in **Fig. 1**) is similar to cholesteryl acetate but with a larger alkyl chain in the ester group. PELE simulations show that the ligand does not approach the heme group from the C25 side (**Fig. S9**) nor by the C3 (data not shown). The reason for this is the large hydrophobic chains that favor interactions with the protein to avoid solvent exposure. An unusual binding mode occurs for this molecule where a strong minimum energy structure is observed that hinders the correct placement of the molecule in the active site by either end.
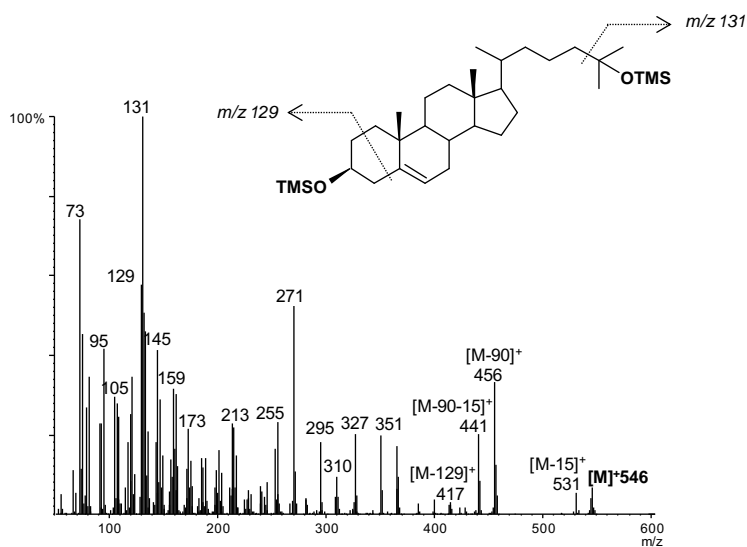
Supplemental Figures

FIG S1 Mass spectrum of 25-hydroxycholesterol from peroxygenase reactions with cholesterol (see **Fig. 1**), as TMS derivative.
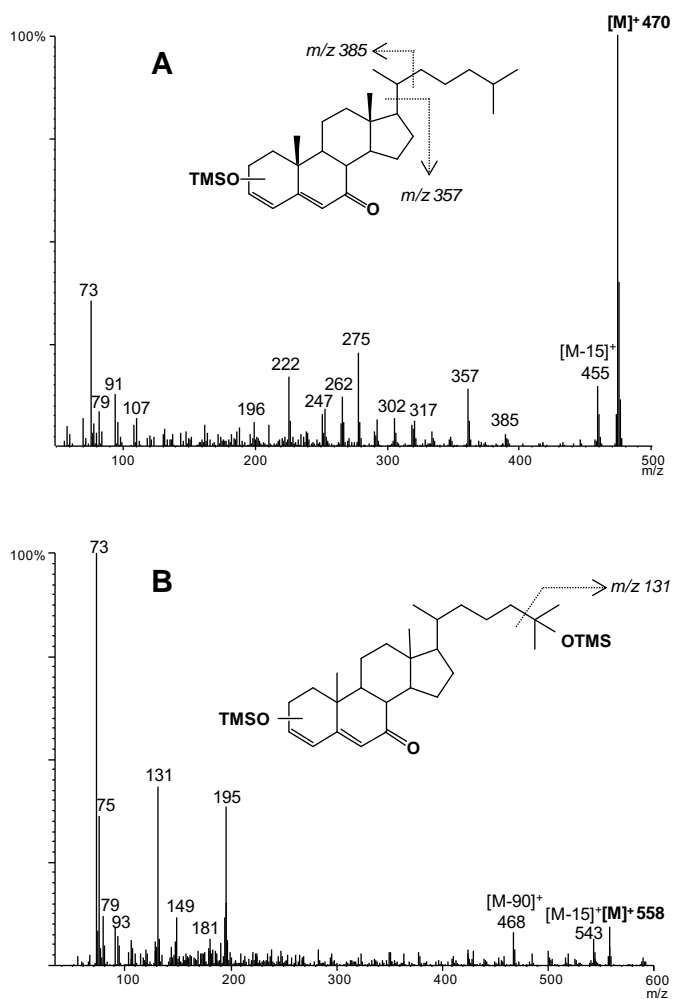
FIG S2 Mass spectra of x-hydroxycholesta-3,5-diene-7-one (**A**) and x,25-dihydroxycholesta-3,5-diene-7-one (**B**) from peroxygenase reactions with cholesta-3,5-diene-7-one (see **Fig. 1**), as TMS derivatives.
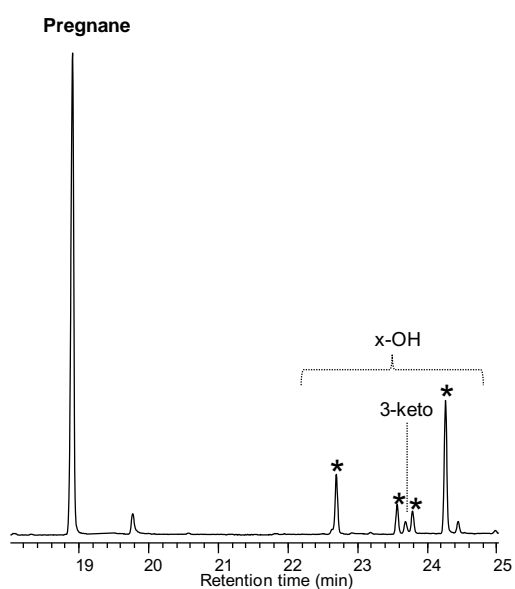
FIG S3 GC-MS analysis of the *Mro*UPO reaction (at 60 min) with pregnane (see **Fig. 1**) showing the remaining substrate and the monohydroxylated derivatives (x denotes the unknown position of hydroxylation in the ring).
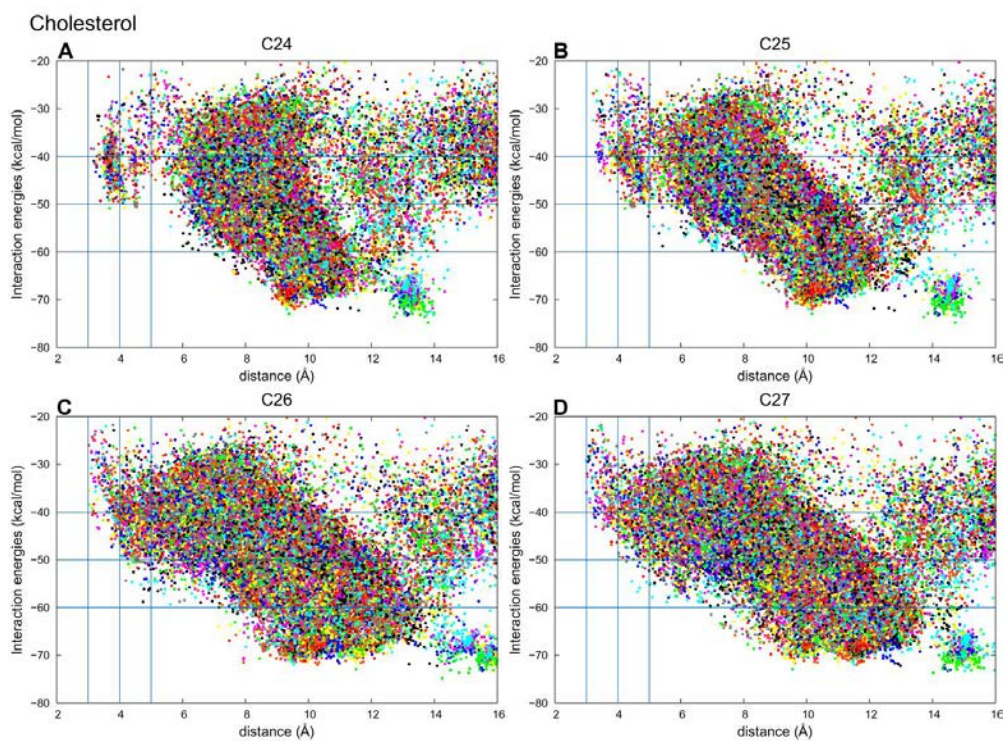
FIG S4 Results from PELE simulations for cholesterol (compound A in **Fig. 1**). The distances plotted correspond to the C24 (**A**), C25 (**B**), C26 (**C**) or C27 (**D**) atoms to the oxygen atom in *Aae*UPO compound I.
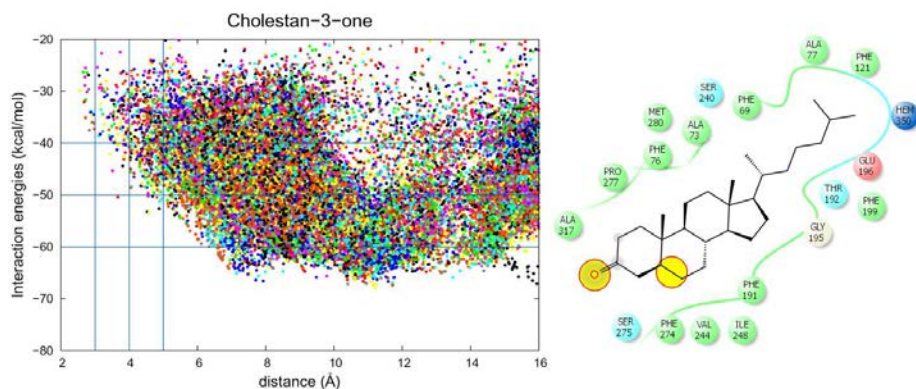
FIG S5 Results from PELE simulations for cholestan-3-one (compound F in **Fig. 1**) diffusion at the active site of the *Aae*UPO including: *(left)* Plots of the energy profile *vs.* the distance between the steroid H25 and the oxygen atom in enzyme compound I; and (*right*) Main interactions between cholestan-3-one and the enzyme in a representative structure in the binding site. The yellow circles identify structural differences relative to cholesterol.
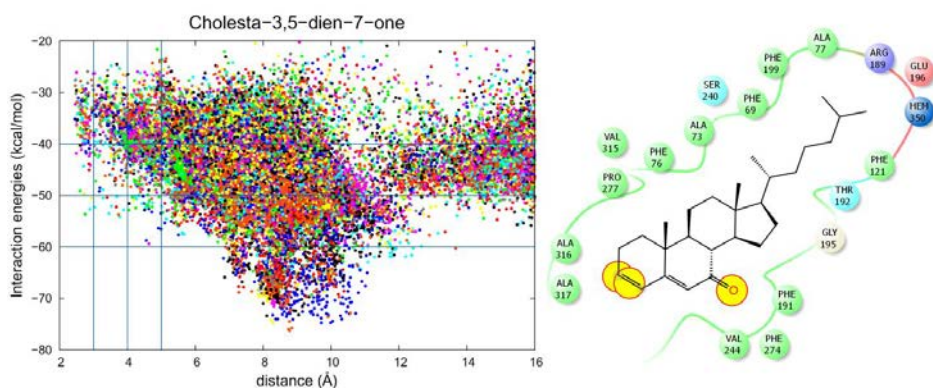

FIG S6 Results from PELE simulations for cholesta-3,5-dien-7-one (compound H in **Fig. 1**) diffusion at the active site with C25 approaching the heme cofactor of the *Aae*UPO, including: (*left*) Plots of the energy profile *vs.* the distance between the steroid H25 atom and the oxygen atom in enzyme compound I; and (*right*) Main interactions between cholesta-3,5-dien-7-one and the enzyme in a representative structure in the binding site. The yellow circles identify structural differences relative to cholesterol.
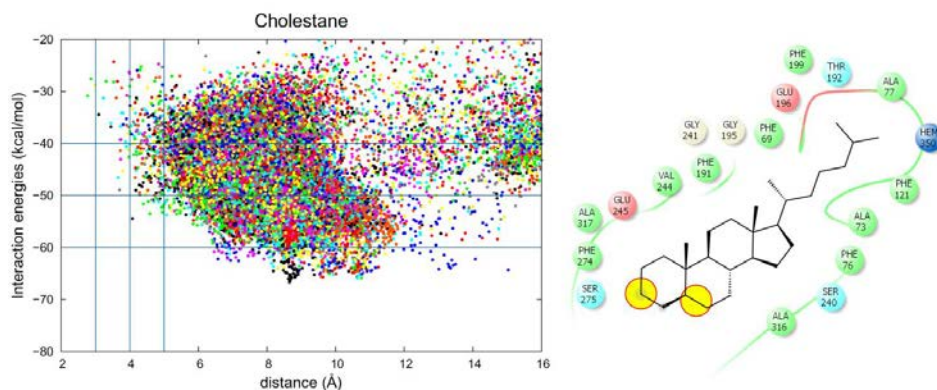
FIG S7 Results from PELE simulations for cholestane (compound J in **Fig. 1**) diffusion at the active site with C25 approaching the heme cofactor of the *Aae*UPO, including: (*left*) Plots of the energy profile *vs* the distance between the steroid C25 atom and the oxygen atom in enzyme cholestane; and (*right*) Main interactions between cholestane and the enzyme in a representative structure in the binding site. The yellow circles identify structural differences relative to cholesterol.
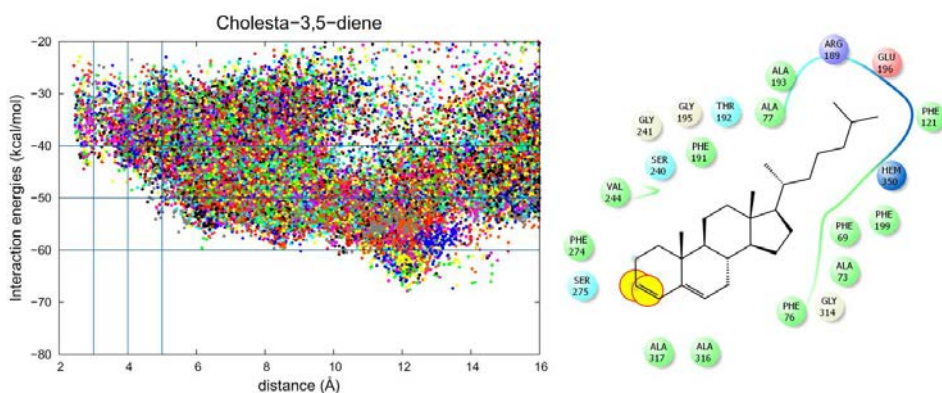

FIG S8 Results from PELE simulations for cholesta-3,4-diene (compound K in **Fig. 1**) diffusion at the active site with C25 approaching the heme cofactor of the *Aae*UPO, including: (*left*) Plots of the energy profile *vs.* the distance between the steroid H25 atom and the oxygen atom in enzyme compound I; and (*right*) Main interactions between cholesta-3,4-diene and the enzyme in a representative structure in the binding. The yellow circles identify structural differences relative to cholesterol.
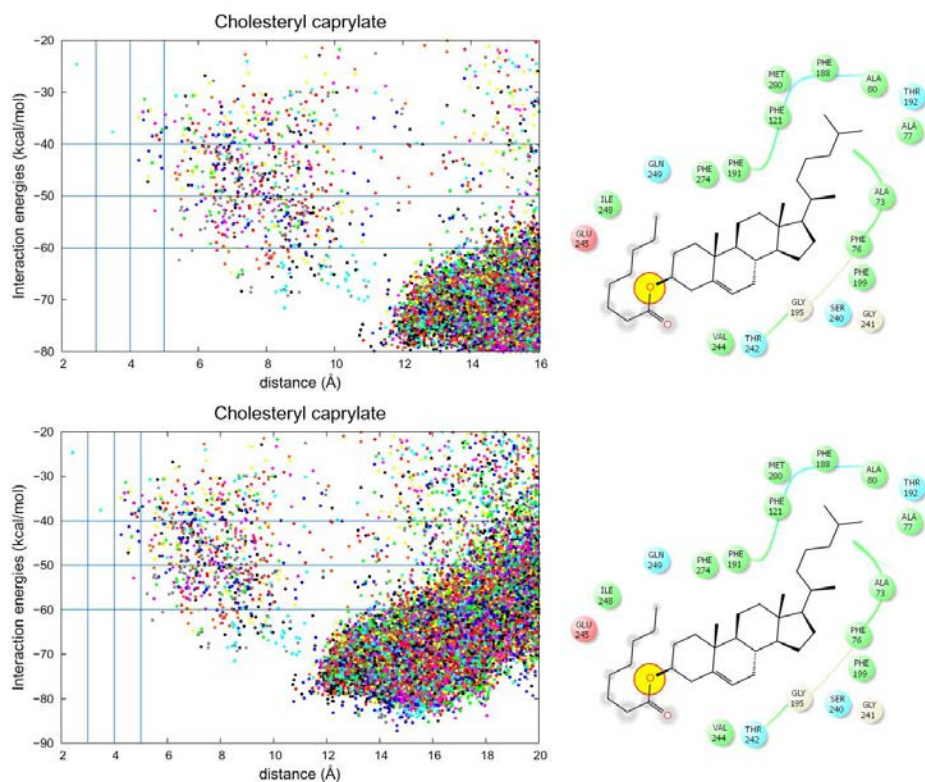
FIG S9 Results from PELE simulations for cholesteryl caprylate (compound O in **Fig. 1**) diffusion at the active site with C25 approaching the heme cofactor of the *Aae*UPO, including: (*left*) Plots of the energy profile *vs.* the distance between the steroid H25 atom and the oxygen atom in enzyme compound I; and (*right*) Main interactions between cholesteryl caprylate and the enzyme in a representative structure in the binding site. The yellow circles identify structural differences relative to cholesterol.

3.1.3 *Article III:* Molecular determinants for selective C_{25}-hydroxylation of vitamins D_2 and D_3 by fungal peroxygenases.

Fátima Lucas,‡*[a] Esteban D. Babot,‡[b] Marina Cañellas,‡[a,c] José C. del Río,[b] Lisbeth Kalum,[d] René Ullrich,[e] Martin Hofrichter,[e] Victor Guallar,[a,f] Angel T. Martínez[g] and Ana Gutiérrez*[b]

[a] Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain. E-mail: **fatima.lucas@bsc.es**

[b] Instituto de Recursos Naturales y Agrobiología de Sevilla, CSIC, Reina Mercedes 10, E-41012 Seville, Spain. E-mail: **anagu@irnase.csic.es**; Fax: +32 954624002; Tel: +32 954624711

[c] Anaxomics Biotech, Balmes 89, E-08008 Barcelona, Spain

[d] Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark

[e] TU Dresden, Department of Bio- and Environmental Sciences, Markt 23, 02763 Zittau, Germany

[f] ICREA, Barcelona, Spain

[g] Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain

‡ These three authors equally contributed to this work

Catalysis Science & Technology

ARTICLE

Molecular determinants for selective C$_{25}$-hydroxylation of vitamins D$_2$ and D$_3$ by fungal peroxygenases

Fátima Lucas,[1,a*] Esteban D. Babot,[1,b] Marina Cañellas,[1,a,c] José C. del Río,[b] Lisbeth Kalum,[d] René Ullrich,[e] Martin Hofrichter,[e] Victor Guallar,[a,f] Angel T. Martínez[g] and Ana Gutiérrez[b*]

Hydroxylation of vitamin D by *Agrocybe aegerita* and *Coprinopsis cinerea* peroxygenases was investigated in a combined experimental and computational study. 25-Monohydroxylated vitamin D$_3$ (cholecalciferol) and D$_2$ (ergocalciferol), compounds of high interest in human health and animal feeding, can be obtained through reaction with both fungal enzymes. Differences in conversion rates and, especially, in site selectivity were nevertheless observed. To rationalize the results, diffusion of D$_2$ and D$_3$ on the molecular structure of the two enzymes was performed with PELE software. In good agreement with experimental conversion yields, simulations indicate more favorable energy profiles for the substrates' entrance in *C. cinerea* than for *A. aegerita* enzyme. On the other hand, GC-MS analyses show that while a full regioselective conversion into the active C$_{25}$ form is catalyzed by *C. cinerea* peroxygenase for D$_2$ and D$_3$, *A. aegerita* yielded a mixture of the hydroxylated D$_3$ products. From the molecular simulations, relative distance distributions between the haem compound I oxygen and H$_{24}$/H$_{25}$ atoms (hydrogens on C$_{24}$ and C$_{25}$ respectively) were plotted. Results show large populations for O-H$_{25}$ distances below 3 Å for D$_2$ and D$_3$ in *C. cinerea* in accordance with the high reactivity observed for this enzyme. In *A. aegerita*, however, cholecalciferol has similar populations (below 3 Å) for O-H$_{25}$ and O-H$_{24}$ which can justify the hydroxylation observed in C$_{24}$. In the case of ergocalciferol, due to the bulky methyl group in position C$_{24}$, very few structures are found with O-H$_{24}$ distances below 3 Å and thus, as expected, reaction was only observed at C$_{25}$ position.

Introduction

Selective oxygenations of aliphatic compounds are among the most challenging reactions in organic chemistry for the regio and/or stereo specific synthesis of pharmaceuticals and fine chemicals. Monooxygenases catalyzing such hydroxylation reactions include cytochromes P450, a family of haem proteins playing a variety of physiological roles but often requiring an auxiliary flavoenzyme (or flavin-containing module) and a source of reducing power to be activated by O$_2$, two facts that limit their biotechnological applicability.

Recently, a new peroxidase type, which shares the active-site architecture and reaction mechanism of cytochromes P450, but has the advantage of being activated directly by H$_2$O$_2$, was isolated from *Agrocybe aegerita*,[46] and later identified in a variety of sequenced basidiomycete genomes including that of *Coprinopsis cinerea*.[14] Due to the above characteristics, these unspecific peroxygenases (EC 1.11.2.1) have a huge biotechnological potential as self-sufficient monooxygenases,[10,20] for hydroxylation of both aromatic[2,3,25-28,43-45] and aliphatic compounds.[4,16,34]

The *A. aegerita* enzyme has been the most widely investigated basidiomycete peroxygenase, but recent studies have shown that the *C. cinerea* enzyme has comparative advantages related to its high conversion yield/selectivity for some hydroxylation reactions, and its production as a recombinant protein in an industrial expression host (by Novozymes, Bagsvaerd, Denmark).[5,7]

Hydroxylation of vitamin D for the selective production of its active C$_{25}$-hydroxylated derivatives is one of the reactions where the *C. cinerea* peroxygenase can be of biotechnological interest (**Fig. 1**).[6] Supplementation with 25-hydroxyvitamin D has a positive effect in different human diseases,[12,21,22,29,42] and also raises considerable interest for feeding broiler chickens[15,18,24,33] and other farm animals,[41] to reduce skeleton problems caused by rapid growth and reduced mobility. Therefore, the use of a peroxygenase in vitamin D hydroxylation represents an attractive alternative to the chemical synthesis.

The goal of the present work was to rationalize the differences observed in cholecalciferol and ergocalciferol (vitamins D$_3$ and D$_2$, respectively) conversion rates and site selectivity by the *A. aegerita* and *C. cinerea* peroxygenases. The work presented here consists first of the enzymatic conversion of these compounds by the two peroxygenases, under the same reaction conditions, and gas chromatography-mass spectrometry (GC-MS) analyses to identify all reaction products. Then, energy profiles and binding modes of vitamins D$_2$ and D$_3$ were determined by structure-based computational simulations, using the PELE software.[11,13] Finally, differences in site selectivity were investigated through the analysis

Catalysis Science & Technology

of the most favorable binding orientations in the active site. Results show that molecular simulations can effectively discriminate experimentally observed differences in conversion rates and site selectivity.
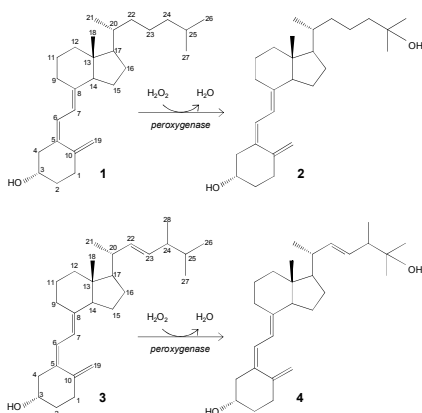

Fig. 1 Enzymatic conversion of cholecalciferol (vitamin D_3; **1**) and ergocalciferol (vitamin D_2; **3**) into their bioactive 25-hydroxylated derivatives (**2** and **4**) by basidiomycete peroxygenases.

Results and discussion

Experimental hydroxylation reactions

Conversion of cholecalciferol and ergocalciferol was experimentally determined for both the *A. aegerita* and *C. cinerea* peroxygenases. In the case of cholecalciferol (0.1 mM), GC-MS analyses of the reaction mixture revealed that this compound was completely (100%) converted by the *C. cinerea* enzyme within 60 min reaction (**Fig. 2C**) as compared with the control reaction without peroxygenase (**Fig. 2A**). In the *A. aegerita* peroxygenase reaction up to 90% conversion was observed (**Fig. 2B**). In the case of ergocalciferol (**Figs. 2D-F**), a conversion of 85% was produced by *C. cinerea* while in *A. aegerita* reaction 81% product was observed. When higher substrate concentrations were tested (e.g. 0.5 mM) higher differences in the conversion rates by the two enzymes were observed (40-50% and 20% conversions by the *C. cinerea* and *A. aegerita* peroxygenases, respectively).

Moreover, cholecalciferol and ergocalciferol conversion by the *C. cinerea* peroxygenase showed a strict site selectivity since it gave exclusively 25-hydroxycalciferol. Likewise, conversion by the *A. aegerita* enzyme of ergocalciferol yielded exclusively 25-hydroxyergocalciferol but for cholecalciferol a mixture of products was observed. The products include 25-hydroxycholecalciferol (64% of the initial substrate) together with 24-hydroxycholecalciferol (21%) and 26/27-hydroxycholecalciferol (7%), as confirmed by comparison with true standards showing identical retention times and mass spectra. The double peaks observed in the chromatograms for both substrate and product (**Fig. 2**) correspond to the isopyro (19β, 9β) and pyro (19α, 9α) isomers formed by thermal rearrangement involving ring-B closure, that vitamin D and its hydroxylated derivatives undergo due to the temperature at which GC-MS EI(+) is carried out. Indeed, the presence of the

two isomers during GC separation is a useful indication that a secosteroid of the vitamin D type was injected into the system.[32] The position of the hydroxyl group at the target C_{25} position was established by MS of the TMS derivative. The spectrum shows a prominent ion from C_{24}-C_{25} bond cleavage, with characteristic fragment at *m/z* 131 and molecular ion at *m/z* 544. Additionally, characteristic fragments at *m/z* 529 ([M-15]$^+$), *m/z* 454 ([M-90]$^+$), *m/z* 439 ([M-90-15]$^+$), *m/z* 349 ([M-90-90-15]$^+$) and *m/z* 413, were also present. Therefore, both the chromatographic profiles and the mass spectra correspond to the isomerized secosteroids.
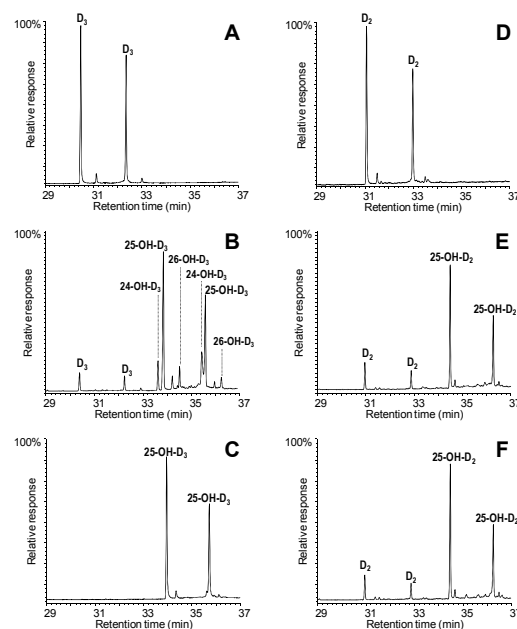

Fig. 2 GC-MS analyses of cholecalciferol (left) and ergocalciferol (right) hydroxylation by the *A. aegerita* (**B, E**) and *C. cinerea* peroxygenases (**C, F**), compared with a control without enzyme showing the substrate peaks (**A, D**), as TMS derivatives from 60-min reactions. In all cases the isopyro (left) and pyro (right) isomers from secosteroid thermal rearrangement are obtained (two small peaks in **A** and **D** correspond to minor additional isomers).

Peroxygenase structure

A superimposition of the general structure and the haem pocket residues in the *A. aegerita* and *C. cinerea* peroxygenases is shown in **Fig. 3**. The main differences are two longer loops (**Fig. 3A**, arrows) and the substitution of Phe69 by Met69 (**Fig. 3B**, red label) in the *C. cinerea* enzyme. All other haem pocket residues, including the proximal cysteine acting as the fifth ligand of the haem iron (Cys36) and the distal glutamic acid and arginine involved in haem activation by H_2O_2 (Glu169 and Arg189),[35] are conserved in the two enzymes. Moreover, differences in the channel providing access to the haem cofactor and the neighbor residues at the channel entrance are shown in **Fig. 4**.
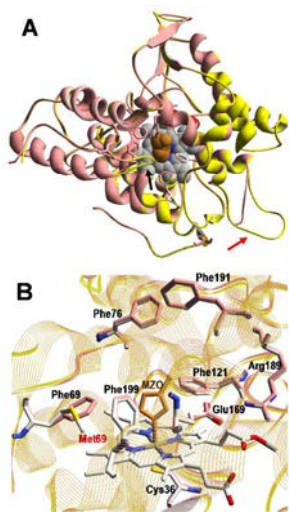
Catalysis Science & Technology

Fig. 3 Comparison of peroxygenase molecular structures. **A)** General superimposition with the *A. aegerita* and *C. cinerea* proteins as pink and yellow ribbons, respectively. The haem group (CPK-colored spheres) and several haem pocket residues (pink- and CPK-colored sticks, respectively) are shown, and two larger loops in the second enzyme are indicated by arrows. **B)** Haem pocket residues in the *A. aegerita* (pink sticks) and *C. cinerea* (CPK-colored sticks) peroxygenases (red label indicates non-conserved methionine in position 69 of the latter enzyme). A hypothetical product molecule identified as 4-hydroxymethylimidazole is shown in **A** (brown-colored vdW spheres) and **B** (MZ0, brown-colored sticks).[35] From *A. aegerita* PDB 2YOR, and *C. cinerea* homology model (provided as **Supplemental file 1**).

Computational modeling

Ligand diffusion energy profiles

For the simulations of cholecalciferol and ergocalciferol access to the H_2O_2-activated haem in the *A. aegerita* and *C. cinerea* peroxygenases, the haem cofactor was modeled as compound I (a FeIV=O, porphyrin cation radical). The substrate was placed close to the entrance of the haem-access channel of the proteins prepared at the optimal pH for peroxygenase activity (pH 7). This initial location was identified using SiteMap,[40] and from there, the ligand was spawned inside the protein by PELE,[11] following the distance between the reactive O atom in the haem compound I and the cholecalciferol/ergocalciferol C_{25} atom (**Fig. 1**).

PELE simulations were done in two stages: first the substrate is perturbed to reduce the C_{25}-O distance and, when this distance is below 5 Å, the substrate is free to explore the active site cavity (at the haem distal side) with a 15 Å restrain. In the first step, the ligand is perturbed with a combination of large and small translations and rotations, ranging from 0.5 to 1.5 Å for translations, and 0.05 to 0.25 radians for rotations. However, during the second stage, translation range is reduced (0.75- .25 Å) to perform a finer active site exploration. The plots shown in **Fig. 5** correspond to three 48 h simulations each with 80 processors, and show the substrate-C_{25} to haem-O distance *vs.* the interaction energy between the protein and the substrate at each of the different poses explored.
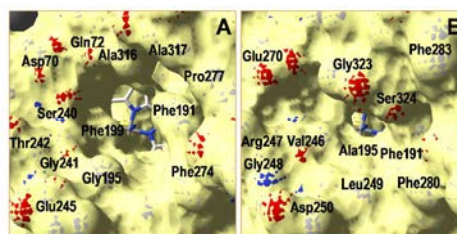

Fig. 4 Solvent access surfaces showing differences in the size of the haem (CPK-colored sticks) access channel in the peroxygenases of *A. aegerita* (**A**) and *C. cinerea* (**B**). Several neighbor residues are shown as CPK-colored vdW spheres, including Gly323 and Ser324 contributing to occlude the haem channel in the *C. cinerea* enzyme.

Simulations show that for the *A. aegerita* peroxygenase the entrance of the substrates from the surface of the protein is quite favorable but then, the access to the activated haem is obtained against an uphill potential (**Figs. 5A,B**). For the *C. cinerea* enzyme the entrance is less open than in the *A. aegerita* peroxygenase and, for this reason, we observed a constrained access to the protein (at around 12 Å). However, the overall energy profile is more favorable, in particular for cholecalciferol (**Fig. 5C**). Once inside the protein, the ligand must surpass smaller barriers to reach the haem. From **Fig. 5** it is clear that *C. cinerea* peroxygenase has the most favorable energy profiles and well defined minima in the active site (with C_{25}-O distance around 3 Å). **Fig. 5A/5C** show that binding of cholecalciferol at the haem site, is more favorable in the *C. cinerea* peroxygenase (-65 kcal/mol) than in the *A. aegerita* enzyme (-40 kcal/mol). This difference comes from the fact that the first protein has a tighter binding pocket (see below). In the case of ergocalciferol **Fig. 5B/5D**, also binding is more favorable in the *C. cinerea* (-42 kcal/mol) than in the *A. aegerita* enzyme (-30 kcal/mol). These differences in the energy profiles, which indicate a more favorable protein-ligand interaction in *C. cinerea,* can explain the higher conversion rate observed for the two compounds in this peroxygenase.
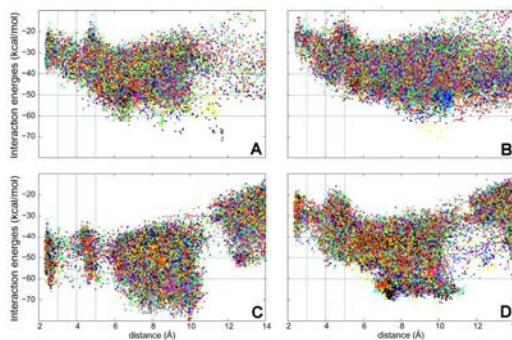

Fig. 5 Interaction energies *vs.* ligand distances from PELE simulations for cholecalciferol (left) and ergocalciferol (right) entrance by the C_{25} end in the peroxygenases from *A. aegerita* (**A, B**) and *C. cinerea* (**C, D**). The distances shown (Å) are between the reactive O atom in the haem compound I and the calciferol C_{25} (for substrate numbering see **Fig. 1**).

Catalysis Science & Technology

Ligand binding

If we overlap cholecalciferol's positions for the entire PELE simulations for both proteins, we find two main orientations in the active site, as shown in **Fig. 6**.
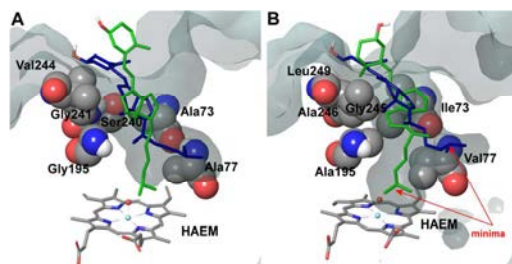

Fig. 6 Superposition of cholecalciferol main active site positions (as liquorice) obtained in PELE simulations on the *A. aegerita* (**A**) and *C. cinerea* (**B**) peroxygenases. Haem cofactor also in liquorice and selected residues as vdW spheres. The protein's active site access is shown as a surface.

The two orientations in **Fig. 6B** correspond to the two minima observed in cholecalciferol's diffusion in the *C. cinerea* peroxygenase (**Fig. 5C**). One of the minima (in green) is found at binding distance to the haem O, and a second (in blue), about 5 Å away from the haem oxygen (**Fig. 5C**). The ligand's positions observed in the *C. cinerea* peroxygenase are also found in the *A. aegerita* enzyme (**Fig. 6A**), although they do not correspond to energy minima (**Fig. 5A**). Likewise, ergocalciferol can adopt two main positions in the binding pocket, although the minima (**Fig. 5D**) are less marked than found for cholecalciferol (**Fig. 5C**).

When cholecalciferol is in an optimal reacting position, it is held in place by *C. cinerea/A. aegerita* peroxygenase Val77/Ala77, Phe121/Phe121, Thr192/Thre192, Ala195/Gly195, Phe199/Phe199 and Glu196/Glu196 (**Fig. S1**). Noteworthy is that Phe69 also contributes to substrate positioning in *A. aegerita* peroxygenase, but the homologous Met69 of the *C. cinerea* enzyme (**Fig. 3B**) is placed away from cholecalciferol and does not appear to have any effect on its position in the active site. Ergocalciferol, however, with an extra methyl group in position C_{24} and with a C_{22}-C_{23} double bond (**Fig. 1**), is positioned in a slightly different manner in the active site, and residues in position 69 (phenylalanine and methionine) now interacts with the ligand (red arrows in **Fig. S2**). This extra constraint could be responsible for the lower reactivity observed for D_2 in both peroxygenases. The influence of side-chain structure (alkyl substituent and unsaturation presence) on the activity of the *A. aegerita* and *C. cinerea* peroxygenases was evidenced in the reaction of these enzymes with different sterols (**Table S1**).

Inspection of the haem entrance when both ligands are in the active site, reveals a better wrapping of the protein around the ligands in *C. cinerea* peroxygenase, compared with the *A. aegerita* enzyme. This is mainly due to the larger loop, where Gly323 is located, which is smaller in the *A. aegerita* peroxygenase (**Fig. 3A**, black arrow). These are better illustrated in the haem-access channel of both peroxygenases in **Fig. 4**, where the position of the above Gly323 is shown. The narrower access to the haem in *C. cinerea* is also the result of several hydrophobic amino acid substitutions. Replacements to larger side chains in *C. cinerea* are dominant with: Ala73/Ile73, Ala77/Val77, Gly195/Ala195, Gly241/Ala246 and Val244/Leu249; and only one to a smaller amino acid, Ser240/Gly245. Due to the larger entrance cavity of the *A. aegerita* peroxygenase, both substrates remained solvent exposed at the C_2

end. In contrast, more favorable interactions are established at the final substrate position inside the tighter channel of the *C. cinerea* peroxygenase, which presents extra interactions on the protein surface. The combination of the surface interactions, along with a tighter haem cavity, result in the improved interaction energies seen for D_2 and D_3 in this protein.

Site selectivity

To investigate the different site selectivities observed for D_2 and D_3 hydroxylation by *A. aegerita* and *C. cinerea* peroxygenases, we haveanalysed the relative distance distribution of the substrates' reactive hydrogen atoms in the active site. We have considered as reactive those that can approach the haem compound I oxygen atom close enough to react. Thus, we have taken into account hydrogen atoms in positions C_{24}, C_{25}, C_{26} and C_{27} for both ligands, in addition to C_{28} hydrogen atoms for ergocalciferol. We have selected all structures (from the PELE simulations) where the distance between H-O is below 5 Å and interaction energies below -20 kcal/mol for D_2 and -30 kcal/mol for D_3. When the relative frequency of these distances was computed (**Fig. 7**) it can be seen that for *C. cinerea* the O-H_{25} frequency is dominant for both compounds. Similar conclusions were obtained for both enzymes and substrates when the distribution of angles Fe=O-H[37] was computed for the different reactive positions in the cholecalciferol and ergocalciferol side-chains (**Fig. S3**).

The percentage of O-H_{25} distances below 3 Å in *C. cinerea* is 54.5% for D_3 and 36.2% for D_2, whereas in *A. aegerita* it is 27.4% and 25.7%, respectively. Moreover, in the D_3 simulations in *A. aegerita*, the fraction of structures with O-H_{24} below 3 Å is 19,3% which is quite high when compared to the other cases. In fact, the fraction of O-H_{25} distances is only 1.4 times superior to O-H_{24} while for the other systems it ranges between 10 to 20 times. This higher fraction of reactive O-H_{24} distances can explain the observed formation of C_{24} hydroxylated products in cholecalciferol. To sum up, and in agreement with experimental results, from the relative frequencies in *C. cinerea* peroxygenase, and also for ergocalciferol in *A. aegerita*, we would expect a completely regioselective reaction in C_{25}. In contrast, for cholecalciferol in *A. aegerita* enzyme, it seems that hydroxyl addition is possible not only in C_{25} position, but also in other carbons (C_{24} and C_{26}/C_{27}).

Conclusions

Atomic level simulations have been used here to rationalize the differences observed for ergocalciferol and cholecalciferol's conversion rates and site selectivity in two peroxygenases. The overall improved energy profiles in *C. cinerea*, the presence of favorable binding, and a high fraction of favorable O-H_{25} distances agrees well with experimentally higher conversion rates.

The main structural differences between the *A. aegerita* and *C. cinerea* peroxygenases that modify the access of cholecalciferol to the activated haem are Ala73/Ile73, Ala77/Val77, Gly195/Ala195, Gly241/Ala246 and Val244/Leu249 changes, all larger amino acids with only Ser240/Gly245 to a smaller one. These larger hydrophobic side chains augment the interaction of D_2 and D_3 substrate with *C. cinerea*. The better conversion rates observed for *C. cinerea* peroxygenase do not originate in the active site itself (where Phe69 is replaced by Met69), but instead in the ligand access to the active site, and especially in the entrance to the protein. In particular, the larger loop hosting Gly323 creates a barrier that reduces the size of the entrance channel in the *C. cinerea* peroxygenase. This, along with a tighter cavity, is reflected in the more favorable interaction energies.
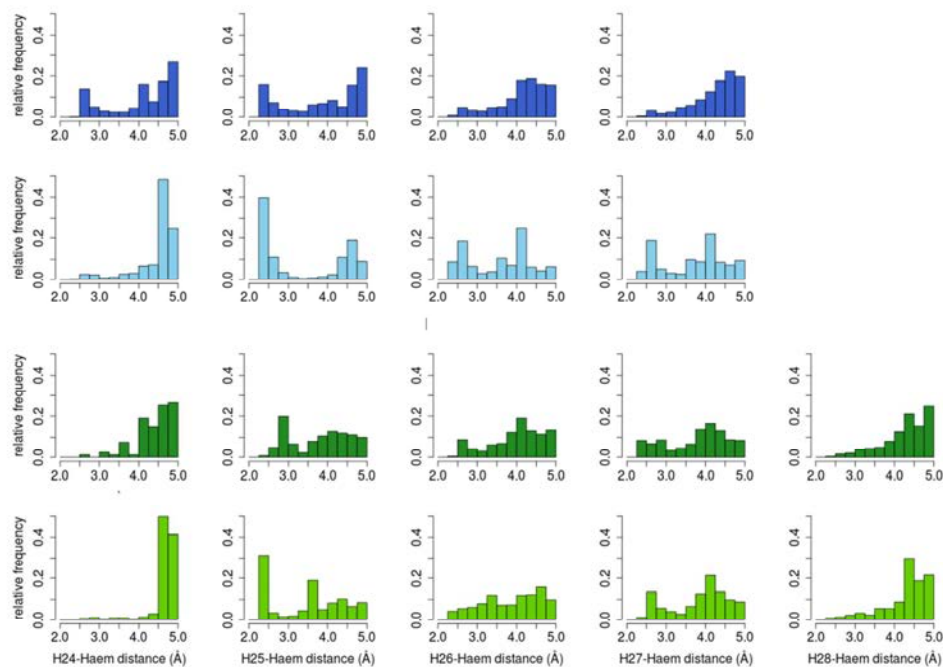
Catalysis Science & Technology

Fig. 7. Relative distance distributions of cholecalciferol (blue) and ergocalciferol (green) reactive hydrogen atoms to haem compound-I oxygen. Histograms for *A. aegerita* are in dark colours, and in light colours for *C. cinerea*.

Although the improved reactivity of *C. cinerea* does not appear to be affected by the larger Phe69 side chain, it does hinder the entrance of ergocalciferol relative to cholecalciferol. The presence of an extra methyl group that interacts directly with position 69 in both proteins lowers the reactivity of this compound. Finally, computed relative frequency of distances between the activated haem oxygen and the hydrogen atoms in C_{24} and C_{25} show that, despite the more favourable hydroxylation on tertiary carbons, reaction at cholecalciferol C_{24} occurs when the ratio of favourable O-H_{25}/O-H_{24} distances decreases.

Materials and methods

Enzymes and chemicals

Peroxygenase (isoform II) was isolated from *A. aegerita* DSM 22459 grown in soybean medium using a combination of SP-Sepharose chromatography and Mono-P chromatofocusing.[46] *A. aegerita* DSM 22459 is deposited at the *Deutsche Stammsammlung für Mikroorganismen und Zellkulturen* Braunschweig (Germany). *C. cinerea* peroxygenase was provided by Novozymes A/S (Bagsvaerd, Denmark). The enzyme corresponds to the protein model 7249 from the sequenced *C. cinerea* genome available at the JGI (http://genome.jgi.doe.gov/Copci1), which was expressed in *Aspergillus oryzae* and purified using a combination of S-Sepharose and SP-Sepharose ion-exchange chromatography (patent

WO/2008/119780). One peroxygenase unit is defined as the amount of enzyme oxidizing 1 μmol of veratryl alcohol to veratraldehyde (ε_{310} 9300 $M^{-1} \cdot cm^{-1}$) in 1 min at 24 ºC, pH 7, after addition of 0.5 mM H_2O_2 in the *C. cinerea* peroxygenase reactions and 2.5 mM H_2O_2 in those with the *A. aegerita* peroxygenase.

Vitamin D_3, also known as calciol or cholecalciferol ((5*Z*,7*E*)-(3*S*)-9,10-seco-5,7,10(19)-cholestatrien-3-ol; **Fig. 1**, structure **1**), was tested as substrate of the *A. aegerita* and *C. cinerea* peroxygenases. 25-Hydroxyvitamin D_3, also known as calcidiol or 25-hydroxycholecalciferol ((5*Z*,7*E*)-(3*S*)-9,10-seco-5,7,10(19)-cholestatriene-3,25-diol; **Fig. 1** structure **2**), was used as standard for gas chromatography-mass spectrometry (GC-MS) analyses. Vitamin D_2, also known as ercalciol or ergocalciferol ((5*Z*,7*E*,22*E*)-(3*S*)-9,10-seco-5,7,10(19),22-ergostatetraen-3-ol; **Fig. 1**, structure **3**), was also tested as substrate of the two peroxygenases. 25-Hydroxyvitamin D_2, also known as ercalcidiol or 25-hydroxyergocalciferol ((5*Z*,7*E*,22*E*)-(3*S*)-9,10-seco-5,7,10(19),22-ergostatetraene-3,25-diol; **Fig. 1**, structure **4**), was also used as standard for GC-MS analyses. All compounds were from Sigma-Aldrich.

Enzymatic reactions

Reactions of cholecalciferol and ergocalciferol (0.1 and 0.5 mM) with the *A. aegerita* and *C. cinerea* peroxygenases (1 U) were performed in 5 mL of 50 mM sodium phosphate, pH 7, at 40 ºC for 60 min, in the presence of 0.5 mM H_2O_2. The substrates were

Catalysis Science & Technology

previously dissolved in acetone, and added to the buffer (the acetone concentration in the reaction was 40%). In control experiments, the substrates were treated under the same conditions (including H_2O_2) but without enzyme. After the enzymatic reactions, products were recovered by liquid-liquid extraction with methyl *tert*-butyl ether, dried under N_2, and redissolved in chloroform for GC-MS analyses. Bis(trimethylsilyl)trifluoroacetamide (Supelco) in the presence of pyridine was used to prepare trimethylsilyl (TMS) derivatives. An internal standard was added after the enzymatic reactions to determine product yields.

GC-MS analyses

GC-MS analyses were performed with a Shimadzu QP2010 Ultra equipment, using a fused-silica DB-5HT capillary column (30 m x 0.25 mm internal diameter, 0.1 μm film thickness) from J&W Scientific.[17] The oven was heated from 120 °C (1 min) to 300 °C (15 min) at 5 °C·min⁻¹. The injection was performed at 300 °C, the transfer line was kept at 300 °C, and helium was used as carrier gas. Compounds were identified by mass fragmentography, and by comparing their mass spectra with standards, and quantitation was obtained from total-ion peak area, using molar response factors obtained from cholecalciferol, ergocalciferol, 25-hydroxycholecalciferol and 25-hydroxyergocalciferol standards. The two latter compounds were also used as external standards for calculation of product yields.

PELE and other computational analyses

The starting structures for PELE simulations were the *A. aegerita* peroxygenase crystal (2YOR)[35] and a homology model for the *C. cinerea* peroxygenase structure obtained using 2YOR as template.[9] As the optimum pH for peroxygenase activity is 7, the structures were prepared accordingly using Schrodinger's Protein Preparation Wizard[36] and H++ web server.[1] Histidines were δ-protonated, with the exception of His82 (ε-protonated) and His118 and His251 (double protonated). All acidic residues were deprotonated, except Asp85 that was kept in its protonated state. The ergocalciferol and cholecalciferol molecules were optimized with Jaguar[38] at the DFT/M06 level with the 6-31G** basis and a PBF implicit solvent in order to obtain their electrostatic potential atomic charges. Finally, the haem site was modeled as thiolate-ligated compound I after being fully optimized in the protein environment with quantum mechanics/molecular mechanics (QM/MM) using QSite.[39] The electronic calculations show three unpaired electrons: two located on the oxoiron group and a third on the heme and less than 1% spin contamination.

Once the initial protein structure was prepared and ligands optimized, these were placed manually in identical positions at the entrance of the haem-access channel and PELE simulations were performed.[11] PELE is a Monte Carlo based algorithm that produces new configurations through a sequential ligand and protein perturbation, side chain prediction and minimization steps, freely available at https://pele.bsc.es. New configurations are then filtered with a Metropolis acceptance test, where the energy is described with an all-atom OPLS force field[23] and a surface generalized Born solvent[8]. In this way it is possible to locate and characterize local and global minima structures for the most favorable protein-ligand interactions. PELE has been successfully used in a number of ligand migration studies with both small and large substrates[19,30,31].

Distance and angle distributions were computed after screening all structures from PELE simulations by interaction energy and distance to the compound I oxygen. Structures that showed interaction energies below -2 kcal/mol for ergocalciferol and -3 kcal/mol for cholecalciferol were selected. Reactive hydrogen atoms

where considered below 5 Å to the compound I oxygen atom for the distance analysis and 3Å for the angle study (angles are much more sensitive and thus only near attack conformations were selected). All the selected structures where then used to compute the relative frequencies of the ligand's hydrogen atoms at reactive distances to the haem oxygen and Fe=O-H and O-H-C angles.

Notes and references

[1] These three authors equally contributed to this work
[a] Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain. E-mail: fatima.lucas@bsc.es
[b] Instituto de Recursos Naturales y Agrobiología de Sevilla, CSIC, Reina Mercedes 10, E-41012 Seville, Spain. E-mail: anagu@irnase.csic.es; Fax: +32 954624002 ; Tel: +32 954624711
[c] Anaxomics Biotech, Balmes 89, E-08008 Barcelona, Spain
[d] Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark
[e] TU Dresden, Department of Bio- and Environmental Sciences, Markt 23, 02763 Zittau, Germany
[f] ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain
[g] Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain

Electronic Supplementary Information (ESI) available: Molecular model for *C. cinerea* peroxygenase obtained using the *A. aegerita* crystal structure as template. See DOI: 10.1039/b000000x/

1. R. Anandakrishnan, B. Aguilar and A. V. Onufriev, *Nucleic Acids Res.*, 2012, **40**, W537-W541.
2. E. Aranda, M. Kinne, M. Kluge, R. Ullrich and M. Hofrichter, *Appl. Microbiol. Biotechnol.*, 2009, **82**, 1057-1066.
3. E. Aranda, R. Ullrich and M. Hofrichter, *Biodegradation*, 2010, **21**, 267-281.
4. E. D. Babot, J. C. del Río, M. Cañellas, F. Sancho, F. Lucas, V. Guallar, L. Kalum, H. Lund, G. Gröbe, K. Scheibner, R. Ullrich, M. Hofrichter, A. T. Martínez and A. Gutiérrez, *Appl. Environ. Microbiol.*, 2015, **81**, 4130-4142.
5. E. D. Babot, J. C. del Río, L. Kalum, A. T. Martínez and A. Gutiérrez, *Biotechnol. Bioeng.*, 2013, **110**, 2332.
6. E. D. Babot, J. C. del Río, L. Kalum, A. T. Martínez and A. Gutiérrez, *Chemcatchem*, 2014, (on line).
7. E. D. Babot, J. C. del Río, L. Kalum, A. T. Martínez and A. Gutiérrez, *Chemcatchem*, 2014, (on line).
8. D. Bashford and D. A. Case, *Annual Review of Physical Chemistry*, 2000, **51**, 129-152.
9. L. Bordoli, F. Kiefer, K. Arnold, P. Benkert, J. Battey and T. Schwede, *Nat. Protoc.*, 2009, **4**, 1-13.
10. S. Bormann, A. G. Baraibar, Y. Ni, D. Holtmann and F. Hollmann, *Catal. Sci. Technol.*, 2015, **5**, 2038-2052.
11. K. W. Borrelli, A. Vitalis, R. Alcantara and V. Guallar, *J. Chem. Theory Comput.*, 2005, **1**, 1304-1311.
12. N. R. Buck, W. Claerhout, B. H. Leuenberger, E. Stoecklin, K. Urban and S. Wolfram, *Patent (USA)*, 2013, US 20130210782 A1.
13. B. P. Cossins, A. Hosseini and V. Guallar, *J. Chem. Theory Comput.*, 2012, **8**, 959-965.
14. D. Floudas, M. Binder, R. Riley, K. Barry, R. A. Blanchette, B. Henrissat, A. T. Martínez, R. Otillar, J. W. Spatafora, J. S. Yadav, A. Aerts, I. Benoit, A. Boyd, A. Carlson, A. Copeland, P. M. Coutinho, R. P. de Vries, P. Ferreira, K. Findley, B. Foster, J. Gaskell, D.

Catalysis Science & Technology

Glotzer, P. Górecki, J. Heitman, C. Hesse, C. Hori, K. Igarashi, J. A. Jurgens, N. Kallen, P. Kersten, A. Kohler, U. Kües, T. K. A. Kumar, A. Kuo, K. LaButti, L. F. Larrondo, E. Lindquist, A. Ling, V. Lombard, S. Lucas, T. Lundell, R. Martin, D. J. McLaughlin, I. Morgenstern, E. Morin, C. Murat, M. Nolan, R. A. Ohm, A. Patyshakuliyeva, A. Rokas, F. J. Ruiz-Dueñas, G. Sabat, A. Salamov, M. Samejima, J. Schmutz, J. C. Slot, F. St.John, J. Stenlid, H. Sun, S. Sun, K. Syed, A. Tsang, A. Wiebenga, D. Young, A. Pisabarro, D. C. Eastwood, F. Martin, D. Cullen, I. V. Grigoriev and D. S. Hibbett, *Science*, 2012, **336**, 1715-1719.

15. C. A. Fritts and P. W. Waldroup, *Journal of Applied Poultry Research*, 2003, **12**, 45-52.
16. A. Gutiérrez, E. D. Babot, R. Ullrich, M. Hofrichter, A. T. Martínez and J. C. del Río, *Arch. Biochem. Biophys.*, 2011, **514**, 33-43.
17. A. Gutiérrez, J. C. del Río, F. J. González-Vila and F. Martín, *J. Chromatogr.*, 1998, **823**, 449-455.
18. J. M. Hernández, *Patent (USA)*, 2013,US 20130137662 A1.
19. A. Hernández-Ortega, F. Lucas, P. Ferreira, M. Medina, V. Guallar and A. T. Martínez, *J. Biol. Chem.*, 2011, **286**, 41105-41114.
20. M. Hofrichter and R. Ullrich, *Curr. Opin. Chem. Biol.*, 2014, **19**, 116-125.
21. G. Jean, J. C. Terrat, T. Vanel, J. M. Hurot, C. Lorriaux, B. Mayor and C. Chazot, *Nephrol. Dial. Transplant.*, 2008, **23**, 3670-3676.
22. G. Jones, *Annu. Rev. Nutr.*, 2013, **33**, 23-44.
23. G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474-6487.
24. S. Käppeli, E. Frohlich, S. G. Gebhardt-Henrich, A. Pfulg, H. Schäublin, R. Zweifel, H. Wiedmer and H. H. Stoffel, *Arch. Geflügelk.*, 2011, **75**, 179-184.
25. A. Karich, M. Kluge, R. Ullrich and M. Hofrichter, *AMB Express*, 2013, **3:5**,
26. M. Kinne, C. Zeisig, R. Ullrich, G. Kayser, K. E. Hammel and M. Hofrichter, *Biochem. Biophys. Res. Commun.*, 2010, **397**, 18-21.
27. M. Kluge, R. Ullrich, C. Dolge, K. Scheibner and M. Hofrichter, *Appl. Microbiol. Biotechnol.*, 2009, **81**, 1071-1076.
28. M. Kluge, R. Ullrich, K. Scheibner and M. Hofrichter, *Green Chem.*, 2012, **14**, 440-446.
29. G. A. Leichtmann, J. M. Bengoa, M. J. G. Bolt and M. D. Sitrin, *Amer. J. Clin. Nutr.*, 1991, **54**, 548-552.
30. D. Linde, R. Pogni, M. Cañellas, F. Lucas, V. Guallar, M. C. Baratto, A. Sinicropi, V. Sáez-Jiménez, C. Coscolín, A. Romero, F. J. Medrano, F. J. Ruiz-Dueñas and A. T. Martínez, *Biochem. J.*, 2014, on-line doi:10.1042/BJ20141211,
31. M. F. Lucas and V. Guallar, *Biophys. J.*, 2012, **102**, 887-896.
32. Makin, H. L. J. and D. B. Gower. Steroid analysis. Springer, NY, 2010.
33. M. Michalczuk, D. Pietrzak, J. Niemiec and J. Mroczk, *Pol. J. Food Nutr. Sci.*, 2010, **60**, 121-126.
34. S. Peter, M. Kinne, X. Wang, R. Ulrich, G. Kayser, J. T. Groves and M. Hofrichter, *FEBS J.*, 2011, **278**, 3667-3675.
35. K. Piontek, E. Strittmatter, R. Ullrich, G. Grobe, M. J. Pecyna, M. Kluge, K. Scheibner, M. Hofrichter and D. A. Plattner, *J. Biol. Chem.*, 2013, **288**, 34767-34776.
36. G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, *Journal of Computer-Aided Molecular Design*, 2013, **27**, 221-234.
37. J. C. Schöneboom, S. Cohen, H. Lin, S. Shaik and W. Thiel, *J. Am. Chem. Soc.*, 2004, **126**, 4017-4034.
38. Schrödinger. Jaguar 7.8. LCC, New York, 2011.
39. Schrödinger. QSite 5.7. LCC, New York, 2011.
40. Schrödinger. SiteMap 2.5. LCC, New York, 2011.
41. C. Simoes-Nunes and G. M. Weber, *Patent (European)*, 2004,EP 1516540 B1.
42. M. D. Sitrin and J. M. Bengoa, *Amer. J. Clin. Nutr.*, 1987, **46**, 1011-1015.
43. R. Ullrich, C. Dolge, M. Kluge and M. Hofrichter, *FEBS Lett.*, 2008, **582**, 4100-4106.
44. R. Ullrich and M. Hofrichter, *FEBS Lett.*, 2005, **579**, 6247-6250.
45. R. Ullrich and M. Hofrichter, *Cell. Mol. Life Sci.*, 2007, **64**, 271-293.
46. R. Ullrich, J. Nuske, K. Scheibner, J. Spantzel and M. Hofrichter, *Appl. Environ. Microbiol.*, 2004, **70**, 4575-4581.

Supporting Information

Molecular determinants for selective C$_{25}$-hydroxylation of vitamins D$_2$ and D$_3$ by fungal peroxygenases

Fátima Lucas,[1,a*] Esteban D. Babot,[1,b] Marina Cañellas,[1,a,c] José C. del Río,[b] Lisbeth Kalum,[d] René Ullrich,[e] Martin Hofrichter,[e] Victor Guallar,[a,f] Angel T. Martínez[g] and Ana Gutiérrez[b*]

[1] *These three authors equally contributed to this work*
[a] *Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain, E-mail: fati.lucas@gmail.com*
[b] *Instituto de Recursos Naturales y Agrobiología de Sevilla, CSIC, Reina Mercedes 10, E-41012 Seville, Spain. E-mail: anagu@irnase.csic.es; Fax: +32 954624002; Tel: +32 954624711*
[c] *Anaxomics Biotech, Balmes 89, E-08008 Barcelona, Spain*
[d] *Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark*
[e] *TU Dresden, Department of Bio- and Environmental Sciences, Markt 23, 02763 Zittau, Germany*
[f] *ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain*
[g] *Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain*

The Supporting Information shows the main interactions of cholecalciferol and ergocalciferol at the peroxygenase heme access channel (**Fig. S1** and **S2**, respectively), the effect of side-chain structure on the conversion rates of five sterols by the *A. aegerita* and *C. cinerea* peroxygenases (**Table S1**), and distributions of estimated Fe=O-H and O-H-C angles in the peroxygenase reactions with cholecalciferol and ergocalciferol (**Fig. S3**).
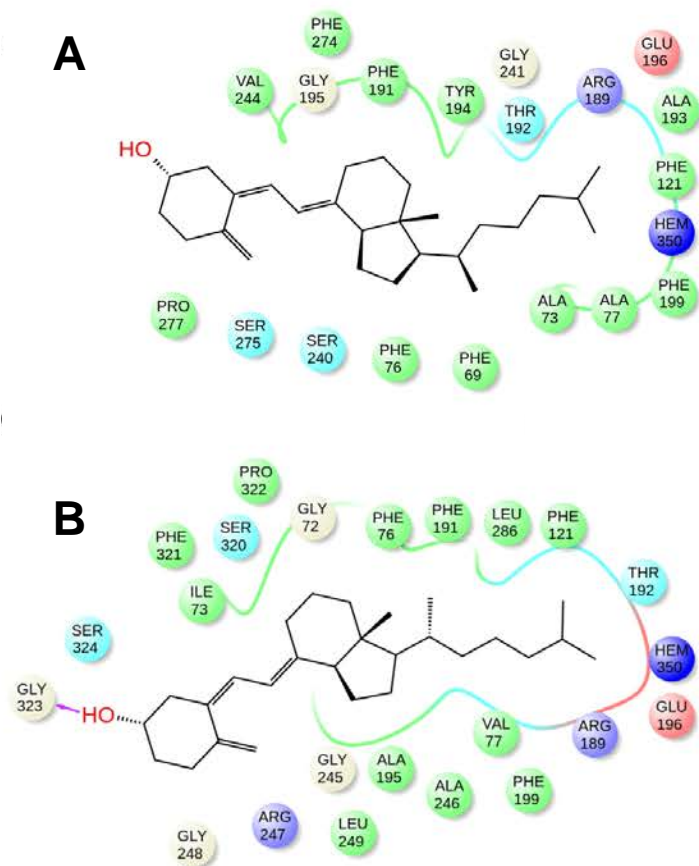
Fig. S1 Main interactions (below 3 Å) for cholecalciferol with the peroxygenases of *A. aegerita* (**A**) and *C. cinerea* (**B**).
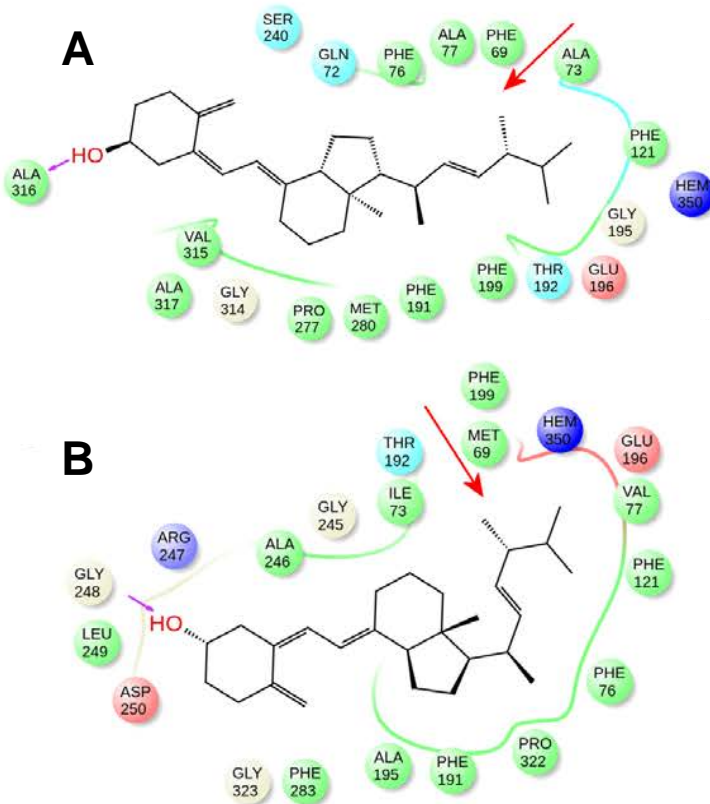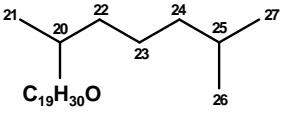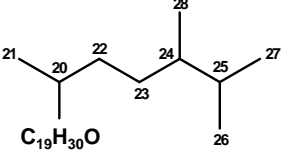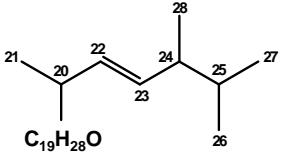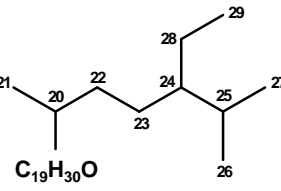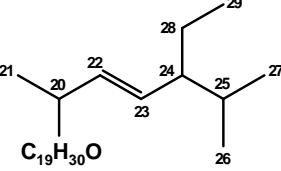
Fig. S2 Main interactions (below 3 Å) for ergocalciferol with the peroxygenases of *A. aegerita* (**A**) and *C. cinerea* (**B**).

Table S1 Conversion degree of steroids with different side-chains (*A-E*: cholesterol, campesterol, ergosterol, sitosterol and stigmasterol, respectively) by *A. aegerita* and *C. cinerea* peroxygenases[a]

		Conversion (%)	
		A. aegerita peroxygenase	*C. cinerea* peroxygenase
A	$C_{19}H_{30}O$	64	100
B	$C_{19}H_{30}O$	47	30
C	$C_{19}H_{28}O$	10	6
D	$C_{19}H_{30}O$	13	6
E	$C_{19}H_{30}O$	2	2

[a]From E. D. Babot, J. C. del Río, M. Cañellas, F. Sancho, F. Lucas, V. Guallar, L. Kalum, H. Lund, G. Gröbe, K. Scheibner, R. Ullrich, M. Hofrichter, A. T. Martínez, and A. Gutiérrez. 2015. Steroid hydroxylation by basidiomycete peroxygenases: A combined experimental and computational study. *Appl. Environ. Microbiol.* 81:4130-4142
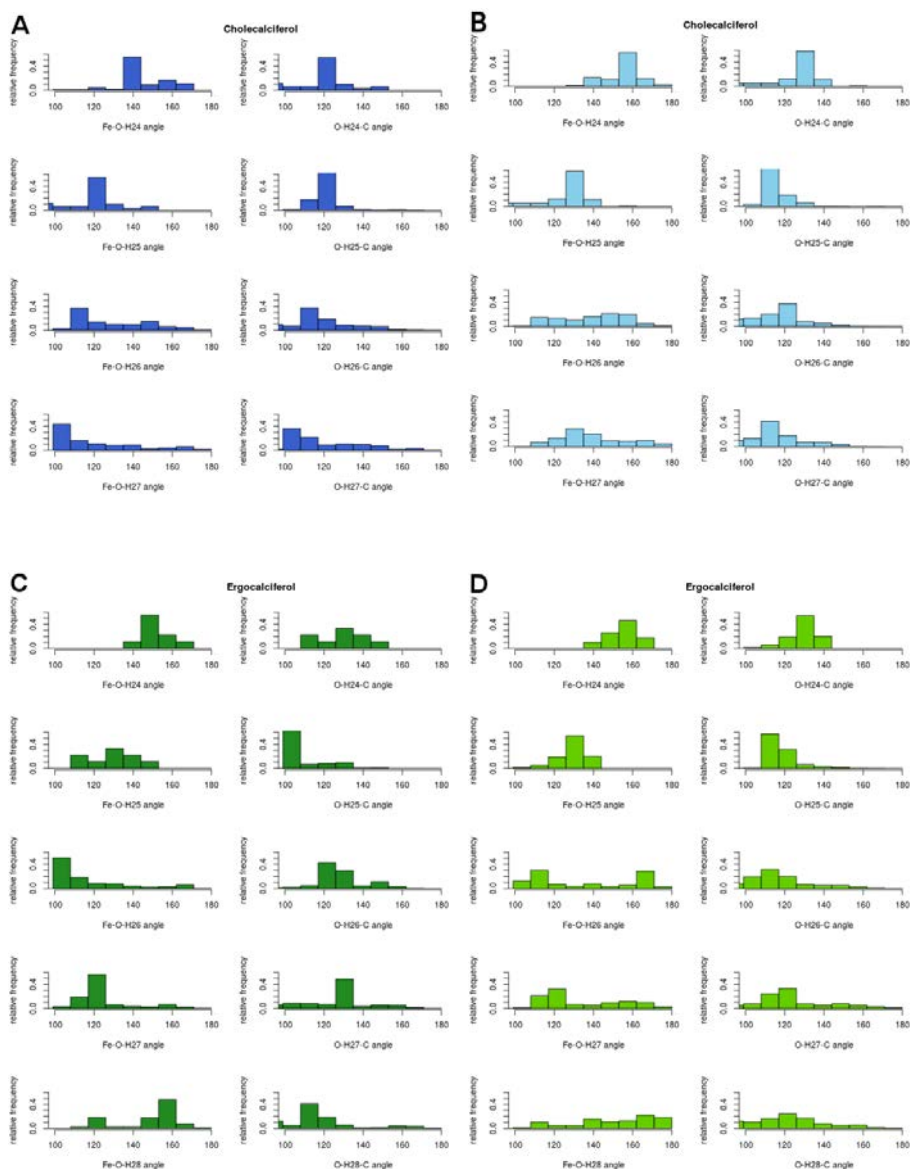
Fig. S3 Relative distributions of computed Fe=O-H and O-H-C angles for the reactions of *A. aegerita* (**A**, **C**) and *C. cinerea* (**B**, **D**) peroxygenases at the C_{24}, C_{25}, C_{26} and C_{27}, positions of cholecalciferol (blue) and ergocalciferol (green) and the C_{28} position of ergocalciferol. Structures were filtered by energy and distance between the pertinent hydrogen atom and the haem compound I oxygen (below 3 Å). QM/MM studies for P450$_{cam}$[b] show that in a pre-arranged reactive position camphor adopts angles of 130° for Fe=O-H and 170° for O-H-C. While for that latter no correlation is found to the theoretical predictions, in the case of Fe=O-H$_{24}$ angles between 120 and 140° are favoured in C_{24} for cholecalciferol reacting with *A. aegerita* (**A**) in

agreement with the experimentally observed formation of 21% product in this position. In the case of the remaining 3 reactions, few structures are found with angles below 145°. In the case of Fe=O-H$_{25}$ (as expected) all compounds present favourable angles. Analogous to the case of the distance distribution analysis, no conclusions can be made for the reactivity of C$_{26}$, C$_{27}$ and C$_{28}$.

[b]J. C. Schöneboom, S. Cohen, H. Lin, S. Shaik and W. Thiel, J. Am. Chem. Soc., 2004, 126, 4017-4034

3.2   ENHANCED VARIANT RATIONALIZATION

Another application of *in silico* approaches is to use them to rationalize and understand the impact that newly introduced mutations have on the catalytic efficiency of enzymes. Using such techniques, one can analyze how a particular mutation modifies the yield of the reaction, the enzyme's chemo-, stereo- or regio-selectivity, the activity ratio when two or more activities are possible (e.g., peroxidase and peroxygenase activities in UPOs), or even the enzyme conformational dynamics [33]. In this way, while QM/MM methods can be used to characterize oxidation and ET processes accurately, biophysical techniques, such as PELE, can be applied to describe substrate migration and recognition [152]. Moreover, these analyses can be used not only to rationalize experimentally produced variants but also to suggest new positions or regions to be mutated, enhancing the success rate of rational mutagenesis. In this section, the application of different computational approaches, along with the knowledge acquired while studying the enzyme's reaction mechanisms in *Articles I-III,* allowed us to rationalize the effect of mutation in different *in vivo* engineered enzymes.

In *Article IV*, the heme pocket of *Aau*DyP was experimentally engineered by our collaborators providing the enzyme with new catalytic sulfoxidase activity. In this way, while the native DyP was unable to perform sulfoxidation reaction, two individual mutations (L357G and F359G) in the heme-binding site provided sulfoxidation activity to DyP with different stereoselectivites. Molecular modelling techniques are here used to unravel the atomic mechanistic details responsible for the observed differences in sulfoxidation, both in yield and selectivity, by the native DyP and its two variants.

On the other hand, UPO enzymes display peroxidative activity (*p*; one-electron oxidations), in addition to its natural peroxygenase activity (*P*; O-transferring two-electron oxidations). From an industrial point of view, the convergence of two activities in the same enzyme could become a problem since the desired products are found along with varying amounts of byproducts of the counterpart activity. In *Articles V* and *VI*, DE is used by our collaborators to tailor *Aae*UPO for the efficient regioselective hydroxylation of high-value chemicals, such as 1-naphthol and 5'-hydroxypropranolol (5'-OHP), obtaining variants with an enhanced *P:p* ratio. In these studies, molecular modelling techniques are used to obtain a detailed atomic explanation of the effects of mutation into enzyme's activity constants, considering both peroxygenase and peroxidase activities. In this way, mutation-induced structural variations could be rationalized and related to changes in the biocatalyst' kinetic constants and regioselectivity.

Publications presented in this section:

Article IV: <u>Cañellas M.,</u> Linde D., Coscolín C., Davó-Siguero I., Romero A., Lucas F., Ruiz-Dueñas F., Guallar V. and Martínez A.T. (2016) Asymmetric sulfoxidation by engineering the heme pocket of a dye-decolorizing peroxidase. *Catalysis Science & Technology.* **6**: 6277-6285.

Article V: Molina-Espeja P., <u>Cañellas M.,</u> Plou F.J., Hofrichter M., Lucas F., Guallar V. and Alcalde M. (2016) Synthesis of 1-Naphthol by a natural peroxygenase engineered by directed evolution. *ChemBioChem.* **17**: 341-349.

Article VI: Gomez de Santos P., <u>Cañellas M.,</u> Tieves F., Younes S.H.H., Molina-Espeja P., Hofrichter M., Hollmann F., Guallar V. and Alcalde M. (2018) Selective synthesis of the human drug metabolite 5'-hydroxypropranolol by an evolved self-sufficient peroxygenase. *ACS Catalysis.* **8**: 4789-4799.

3.2.1 *Article IV:* Asymmetric sulfoxidation by engineering the heme pocket of a dye-decolorizing peroxidase.

Dolores Linde,‡[a] Marina Cañellas,‡[bc] Cristina Coscolín,§[a] Irene Davó-Siguero,[a] Antonio Romero,[a] Fátima Lucas,[bc] Francisco J. Ruiz-Dueñas,[a] Victor Guallar*[bd] and Angel T. Martínez*[a]

[a] Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain. E-mail: **ATMartinez@cib.csic.es**

[b] Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain. E-mail: **victor.guallar@bsc.es**

[c] Anaxomics Biotech, Balmes 89, E-08008 Barcelona, Spain

[d] ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain

‡ These three authors equally contributed to this work.

§ Current address: Instituto de Catálisis y Petreleoquímica, CSIC, Marie Curie 2, 28049 Madrid, Spain.

PAPER

CrossMark
click for updates

Asymmetric sulfoxidation by engineering the heme pocket of a dye-decolorizing peroxidase†

Dolores Linde,‡[a] Marina Cañellas,‡[bc] Cristina Coscolín,§[a] Irene Davó-Siguero,[a] Antonio Romero,[a] Fátima Lucas,[bc] Francisco J. Ruiz-Dueñas,[a] Victor Guallar*[bd] and Angel T. Martínez*[a]

The so-called dye-decolorizing peroxidases (DyPs) constitute a new family of proteins exhibiting remarkable stability. With the aim of providing them new catalytic activities of biotechnological interest, the heme pocket of one of the few DyPs fully characterized to date (from the fungus *Auricularia auricula-judae*) was redesigned based on the crystal structure available, and its potential for asymmetric sulfoxidation was evaluated. Chiral sulfoxides are important targets in organic synthesis and enzyme catalysis, due to a variety of applications. Interestingly, one of the DyP variants, F359G, is highly stereoselective in sulfoxidizing methyl-phenyl sulfide and methyl-*p*-tolyl sulfide (95–99% conversion, with up to 99% excess of the *S* enantiomer in short reaction times), while the parent DyP has no sulfoxidation activity, and the L357G variant produces both *R* and *S* enantiomers. The two variants were crystallized, and their crystal structures were used in molecular simulations to provide a rational explanation for the new catalytic activities. Protein energy landscape exploration (PELE) showed more favorable protein–substrate catalytic complexes for the above variants, with a considerable number of structures near the oxygen atom of the activated heme, which is incorporated into the substrates as shown in ^{18}O-labeling experiments, and improved affinity with respect to the parent enzyme, explaining their sulfoxidation activity. Additional quantum mechanics/molecular mechanics (QM/MM) calculations were performed to elucidate the high stereoselectivity observed for the F359G variant, which correlated with higher reactivity on the substrate molecules adopting pro-*S* poses at the active site. Similar computational analyses can help introduce/improve (stereoselective) sulfoxidation activity in related hemeproteins.

Introduction

There is increasing interest in the synthesis and use of molecules containing stereogenic centers. The sulfur atom of sulfoxides bearing two different substituents is a chiral center, since it adopts tetrahedral sp^3 hybridization, with a lone electron pair occupying the fourth quadrant. Chiral sulfoxides

[a] *Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain. E-mail: ATMartinez@cib.csic.es*
[b] *Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain. E-mail: victor.guallar@bsc.es*
[c] *Anaxomics Biotech, Balmes 89, E-08008 Barcelona, Spain*
[d] *ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain*

have a wide range of applications, from chiral auxiliaries to pharmaceuticals.[1] Among enzymes of interest in chiral synthesis,[2] flavoenzymes (including Baeyer–Villiger monooxygenases) and hemeperoxidases are used for sulfoxidation reactions.[3–5] Cytochrome P450 enzymes also catalyze sulfoxidations, but the requirement of an auxiliary flavin-containing enzyme/domain (and a source of reducing power, as in the case of NAD[P]H-dependent flavoenzymes) limits their biotechnological applicability.[6] The enzymatic production of active *S* omeprazole, a multibillion dollar drug, by a modified cyclohexanone monooxygenase is a good example of these biotransformations.[7]

In peroxidases, the oxidation reaction is mediated by a peroxygenase mechanism rather than a peroxidase mechanism, and sulfoxidation of thioanisole (methyl-phenyl sulfide, MPS) and methyl-*p*-tolyl sulfide (MTS) yielding the corresponding sulfoxides has been used as a probe of oxygen transfer to organic sulfides. In this way, sulfoxidation has been reported for the well-known fungal chloroperoxidase (CPO)[8–10] and horseradish peroxidase (HRP),[10–12] as well as for other haloperoxidases (including vanadium peroxidases),[13] several

Catal. Sci. Technol., 2016, **6**, 6277–6285 | 6277

-132-

animal peroxidases,[14] and also for cytochrome c peroxidase (CcP),[15] *Coprinopsis cinerea* peroxidase (CiP),[16] and lignin peroxidase (LiP),[17] with different conversion rates and stereoselectivities. Recently, a robust peroxidase/peroxygenase of the HTP protein superfamily has been described from the basidiomycete *Agrocybe aegerita* with predominant monooxygenase activity (unspecific peroxygenase, UPO),[18] which is able to selectively convert MPS into the *R* enantiomer of methyl-phenyl sulfoxide.[19,20] Another UPO, from the related fungus *Coprinellus varians*, is especially efficient in sulfoxidizing dibenzothiophene.[21]

Although the natural substrate/s of the so-called dye-decolorizing peroxidases (DyPs), forming part of the new CDE protein superfamily,[22] are still to be identified, these enzymes present considerable catalytic versatility and exceptional stability under extreme pH, temperature, and even pressure conditions.[23,24] DyPs, as some fungal ligninolytic peroxidases of the peroxidase–catalase superfamily (LiP and versatile peroxidase), have the ability to oxidize substrates at two catalytic sites, one located at an exposed protein radical and a second one at the heme pocket, which has a more axial access in DyPs than in ligninolytic and related peroxidases.[25]

In this work, we engineer the heme pocket of the *Auricularia auricula-judae* DyP, one of the few members of this new family fully characterized to date,[26–28] with the aim of providing new catalytic properties to the enzyme. The native recombinant DyP is unable to perform sulfoxidation reaction, but interestingly, two individual mutations at the heme pocket provide sulfoxidation activity on MPS and MTS. More interestingly, one of them was stereoselectively forming the *S* enantiomer, while the second variant yielded similar percentages of both isomers. With the use of molecular modeling techniques, we aim to provide a rational explanation at the atomic level of the differences in yield and selectivity observed during sulfoxidation of MPS and MTS by the native DyP and its two heme pocket variants, whose crystal structures were solved.

Results and discussion

Experimental enzyme engineering and sulfoxidation reactions

The heme pockets of *A. auricula-judae* DyP and other peroxidases were compared, and variants with enlarged access to the enzyme cofactor were obtained and evaluated for sulfoxidation of two organic sulfides, as described in the next three sections.

Redesigning the DyP heme pocket for sulfoxidation. A comparison of the upper side of the heme pocket (where $Fe^{4+}=O$ is located in compound I after H_2O_2 activation) in peroxidases from three different superfamilies is shown in Fig. 1. In HRP (Fig. 1A), the so-called distal histidine occupies an axial position (above the heme iron) contributing to the reaction with H_2O_2 together with a neighbor arginine.[29] In UPO (Fig. 1B), a glutamate/arginine couple plays a similar function,[30] while the couple is aspartate/arginine in DyP (Fig. 1C).[27]
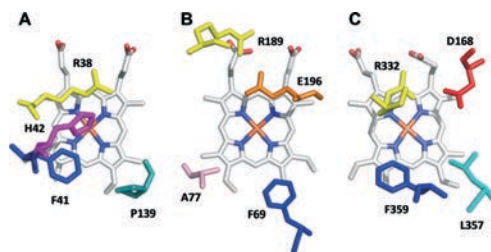

Fig. 1 Upper side of the heme pocket in the crystal structures of (A) HRP (PDB 1ATJ), (B) *A. agerita* UPO (PDB 2YP1), and (C) *A. auricula-judae* DyP (PDB 4W7J) (this heme side is also known as the distal side due to the presence of the distal histidine of peroxidase–catalases, His42 in A).

Due to their crucial role in the formation of reactive compound I, site-directed mutagenesis of the above-mentioned residues in *A. auricula-judae* DyP drastically decreased its catalytic efficiency reducing H_2O_2 (measured with 2,2'-azinobis[3-ethylbenzothiazoline-6-sulfonic acid], ABTS, as the enzyme reducing substrate), which passed from 2050 s^{-1} mM^{-1} in native DyP to only 4 and 1 s^{-1} mM^{-1} in the D168N and R332L variants, respectively. Therefore, other residues contributing to a confined heme pocket in DyP, such as Phe359 (homologous to Phe41 in HRP) and Leu357 (Fig. 1C), were modified, and the enlarged pocket variants were expressed in *E. coli*, *in vitro* activated, purified to homogeneity (with specific activities of 400–500 U mg^{-1} measured with ABTS), and crystallized, and their sulfoxidation activity was evaluated, as described below.

Although the L357G variant showed slightly decreased thermal and pH stabilities (T_{50} = 59 °C and 50% activity at pH 9), the F359G variant did not show significantly modified stability in the range of pH 4–9 (being always higher than 85%) and had only one degree lower T_{50} (61 °C) than the native recombinant DyP (62 °C).

Crystal structures of the DyP L357G and F359G variants. Crystal structures of the L357G (PDB 5IKG) and F359G (PDB 5IKD) variants were solved at 1.9 and 1.1 Å, respectively. Their subsequent analysis confirmed that the overall folding characterizing these enzymes as well as the position of the heme group were conserved. As expected, the only changes observed were in the region where mutations were introduced, and basically consisted in the enlargement of the heme pocket, which was more significant for the F359G variant (Fig. 2A–C), while the channel opening at the protein surface was wider in the L357G variant (Fig. 2D–F).

A close-up view of the heme pocket shows an asymmetric distribution of the cavity according to the volume and position of the mutated residue. Thus, in the F359G variant, the heme pocket is larger allowing the substrate to be properly positioned, with a potential effect on the stereoselectivity of sulfoxidation.

6278 | *Catal. Sci. Technol.*, 2016, **6**, 6277–6285

This journal is © The Royal Society of Chemistry 2016
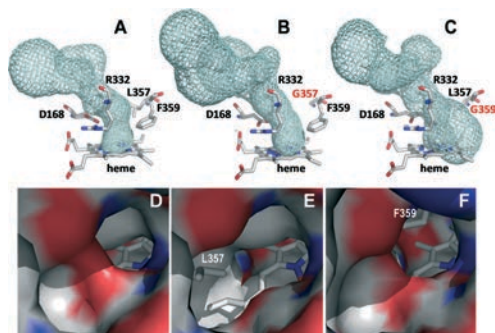
Catalysis Science & Technology

Paper

Fig. 2 Details of the access channel to the heme pocket (top) and its opening at the protein surface (bottom) in the crystal structures of the native DyP (A and D), L357G variant (B and E) and F359G variant (C and F). In A–C, the heme access channel is shown as cyan meshes and the heme and neighbor residues as CPK-colored sticks (the two mutated residues in B and C with red labels), while in D–F an electrostatic potential surface is shown, with the heme as CPK sticks and the removed side chains in the two variants as gray sticks.
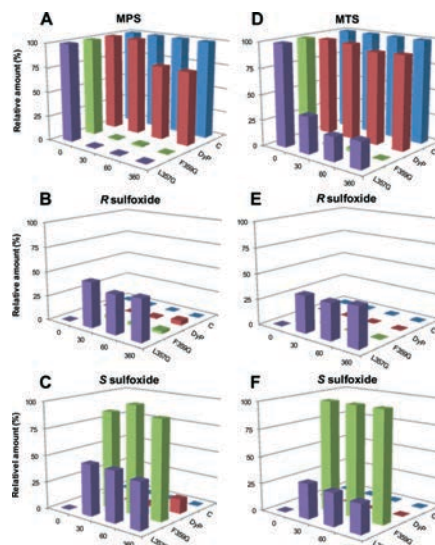

Fig. 3 Results from chiral HPLC analysis of MPS (A–C) and MTS (D–F) reactions with native DyP and L357G and F359G variants and controls (C) without the enzyme (after 0, 30, 60 and 360 min incubation) showing the remaining substrate (A and D) and the resulting R (B and E) and S (C and F) sulfoxides.

Sulfoxidation chromatographic and kinetic analyses. Oxidation of MPS and MTS by the *A. auricula-judae* DyP (native recombinant protein) and its L357G, F359G, F359W and F359H variants was followed in time course reactions using chiral HPLC. Native DyP only produced a small amount of methyl phenyl sulfoxide, 25% MPS conversion after 6 h incubation, and it did not sulfoxidize the bigger substrate MTS (Fig. 3). However, the reactions with the heme pocket variants revealed that L357G and F359G were able to perform sulfoxidation with high efficiency. In the case of MPS, 92% conversion and 95% conversion after 30 min reaction were obtained with the L357G and F359G variants, respectively (Fig. 3A), while 65% conversion and 99% conversion of MTS under the same reaction conditions were obtained, respectively (Fig. 3D). The conversion rates were high (up to 95–99%), and no additional oxidation products (sulfones) were detected. The F359W and F359H variants, including changes that did not enlarge the heme pocket, were unable to sulfoxidize any of the two sulfides assayed. Gas chromatography–mass spectrometry (GC–MS) analyses of reactions using $H_2^{18}O_2$ (90% isotopic labeling) yielded ^{18}O-sulfoxides (88% and 79% labeling for MPS conversion with the F359G and L357G variants, respectively) revealing that sulfoxidation was a peroxygenation reaction (6% ^{18}O-labeling was found in the reactions with $H_2^{16}O_2$ in $H_2^{18}O$ buffer).

These long-term incubation experiments were complemented by estimating kinetic constants for MPS and MTS oxidation under steady-state conditions. F359G showed higher catalytic efficiency (k_{cat}/K_m) than L357G due to 2-fold higher turnover number (k_{cat}) and lower K_m (Table 1). This agrees with the higher catalytic efficiency of the F359G variant oxidizing the standard substrate ABTS at the heme channel (1040 ± 80 s^{-1} mM^{-1}) compared with L357G (175 ± 19 s^{-1}

mM^{-1}). The sulfoxidation catalytic efficiency (and other kinetic constants) of F359G DyP is 2–3 orders of magnitude higher than that of wild-type HRP,[11,12] in the same order of those reported for the best HRP variants,[31–34] and lower than that reported for wild-type CPO (the classic sulfoxidation biocatalysts),[35] and especially for the recently discovered UPO.[20]

Interestingly, in the MPS (Fig. 3B and C) and MTS (Fig. 3E and F) reactions, the enantiomeric production by the two variants is not the same. L357G produces nearly racemic mixtures of the methyl-phenyl and methyl-p-tolyl sulfoxides, while F359G is stereoselectively producing the S sulfoxide. From these values, enantiomeric excesses (ee) of 92% and 99% were calculated for MPS and MTS oxidation by the F359G variant, respectively, while the L357G was scarcely stereoselective, as shown in Table 2. On the other hand, F359G showed a higher total conversion rate than L357G, in agreement with its higher catalytic efficiency shown in Table 1.

Sulfoxidation had been reported for a bacterial (*Thermobifida fusca*) DyP, but the conversion was poor (no rate provided) and the reaction was only slightly selective (61/49% ee after 36 h reaction with MPS/MTS).[36] On the other hand, improvements in sulfoxidation ability have been reported by engineering the heme pocket of HRP,[31–34] and CcP.[15] These included the W51A and F41L variants, in which the bulky tryptophan/phenylalanine side chains in the active sites of CcP and HRP, respectively, were replaced by smaller groups. This agrees with the present results showing that

Table 1 Kinetic constants – k_{cat} (s^{-1}), K_m (mM) and k_{cat}/K_m (s^{-1} mM^{-1}) – for MPS and MTS sulfoxidation with native DyP and two directed variants (means and 95% confidence limits)

	MPS			MTS		
	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m
DyP	0	—	—	0	—	—
L357G	8.0 ± 0.5	0.60 ± 0.08	13.3 ± 1.0	0	—	—
F359G	17.2 ± 0.9	0.37 ± 0.01	45.9 ± 7.4	3.6 ± 0.4	0.13 ± 0.03	26.5 ± 3.8

Table 2 Chiral HPLC analysis of MPS and MTS reactions (30 min) with native DyP and two directed variants

	MPS			MTS		
	Conversion (%)	Isomer	ee (%)	Conversion (%)	Isomer	ee (%)
DyP	0	—	—^{a}	0	—	—
L357G	92	S	2	65	R	8
F359G	95	S	92	99	S	99

^{a} Only 25% conversion of MPS by DyP after 360 min with 42% ee of the S isomer.

enlarging the heme pocket of DyP, as found in the F359G variant, resulted in efficient (and stereoselective) sulfoxidation. Protein engineering has also been reported to improve other sulfoxidation reactions, such as in the Codexis patent for the already mentioned production of S omeprazole by a bacterial monooxygenase.[37]

Computational analyses

For a deeper inquiry on the atomic mechanistic details responsible for the observed differences in sulfoxidation (rate and selectivity) by the native DyP and its two variants, we turned to molecular modeling using the crystal structures solved in this work, as described in the next two sections.

Ligand diffusion energy profiles. When analyzing both the interaction energies and the substrate–heme distances (from the compound **I** oxygen to the sulfur atom of the substrates), protein energy landscape exploration (PELE) simulations show a more favorable protein–substrate catalytic complex for the two variants. In L357G and F359G (Fig. 4B and C, respectively), we find a considerable number of structures below 4 Å (which we consider to be optimal for the reaction) with better interaction energies, with respect to the native enzyme (Fig. 4A).

Moreover, we see a higher number of trajectories at catalytic distances for those conditions where we observe over 90% conversion in 30 min: F359G reaction with MTS and MPS and L357G reaction with MPS. Notice that although in native-DyP MPS interaction energies are quite favorable, the ligand is positioned too far from the heme. These differences in the energy profiles, which indicate an unfavorable ligand–protein interaction and ligand positioning on native DyP, can explain its undetectable (on MTS) or very low (on
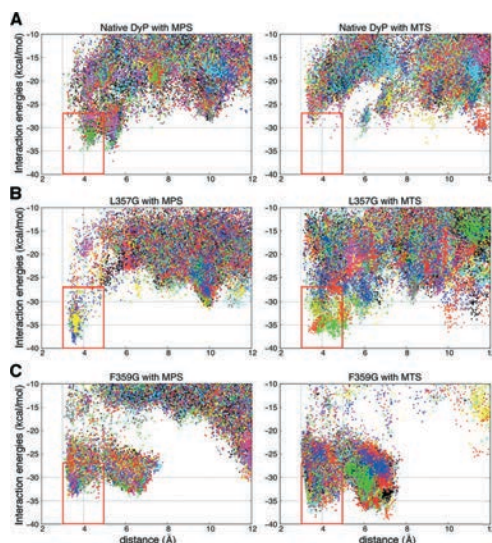

Fig. 4 Interaction energies (in kcal mol^{-1}) vs. ligand distances (in Å) from PELE simulations for MPS (left) and MTS (right) substrates in the (A) native DyP, (B) L357G variant, and (C) F359G variant. The distances are between the reactive O atom in the heme compound **I** and the sulfur (S) atom of the substrates; each color corresponds to a different trajectory. The structures used for the selectivity analysis are delimited with a red box.

MPS) activity, which was only observed in long-term incubation experiments.

Selectivity. To investigate the observed stereoselectivity of the sulfoxidation reaction, a study of the pro-R and pro-S positioning tendency for MPS and MTS was performed. For this, PELE structures from the lowest 10 kcal mol^{-1} interaction energies and with distances to the heme below 5 Å were selected (different criteria to select PELE structures did not significantly change the results). Then, the O–S–C1–C2 dihedral angles were extracted (Fig. 5), and all structures were classified into potential pro-R, pro-S or mixed pro-R/S poses.

First, it should be mentioned that there was a higher number of structures in a good position towards the heme and with an optimal energy to react in the F359G variant than in L357G, in agreement with the higher catalytic
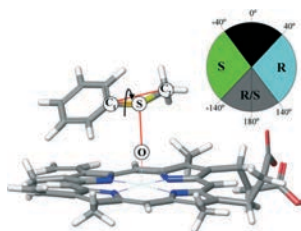
Fig. 5 *R/S* enantiomer classification criterion. The Φ dihedral is computed from the compound **I** catalytic oxygen (O) atom and substrate S and carbon C1 and C2 atoms. When Φ is between 40° and 140°, the substrate is classified as pro-*R*, and when Φ is between −40° and −140°, it is classified as pro-*S*. Otherwise, the substrate is considered to be able to form both *R* and *S* enantiomers equally.

efficiency of the phenylalanine variant. The L357G and, especially, the F359G mutations cause a binding site enlargement, which improves MPS and MTS positioning in comparison with native DyP, as shown by PELE.

Fig. 6 shows the different pro-*R* (cyan) and pro-*S* (green) positions for the native DyP and F359G, the best variant. In DyP, the presence of large (and bulky) residues (Arg332, Leu357 and Phe359) in the heme pocket hinders the proper positioning of both MPS and MTS substrates. Otherwise, for the two mutant proteins, we find an easier positioning of the substrate ring, in correlation with the binding energy analysis shown above. However, contrary to experimental results, substrates on both DyP variants show a preference to adopt pro-*R* positions, even though pro-*S* and pro-*R/S* positions are also possible.

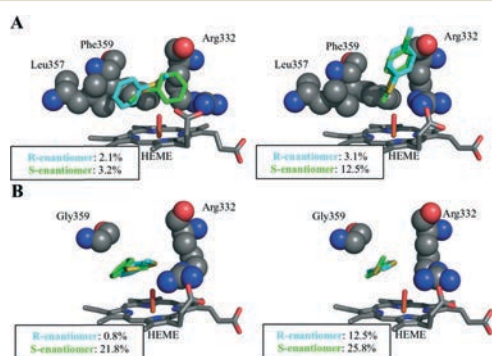To clarify this apparent contradiction with the experimental results, substrate spin densities were computed with

quantum mechanics/molecular mechanics (QM/MM) calculations on the best (substrate) positioned structures. The spin density originates from substrate oxidation, which directly correlates with sulfoxidation.[38,39] In addition to having better protein–ligand interaction energies, as described above, the DyP variants showed a significantly larger spin density than the native DyP, again confirming the improvement in catalytic activity upon mutation. Moreover, in the F359G variant, we observe a substantial difference between the spin density values for the two enantiomers provided in Fig. 6, which might discriminate against the *R* enantiomer oxidation. Thus, even though a higher percentage of structures are placed favoring the formation of an *R* enantiomer, pro-*S* structures are more easily oxidized, which would explain the *S* stereoselectivity experimentally observed for this variant. In agreement with the racemic mixture seen experimentally for the L357G variant, we see a higher relative frequency of the pro-*R/S* position for L357G compared to F359G (Fig. S1A†).

Interestingly, changes in spin density, associated with the different enantiomers, correlate with the position adopted by the substrate with respect to Arg332. Mutations induce a change in the positioning of the ligand that leads to a closer interaction of the pro-*R* structures with this arginine residue (Fig. 7). As seen in previous publications,[40] the electrostatic environment changes caused by this positive charge could lead to an altered substrate oxidation. This effect is larger in the F359G variant, where differences in spin density are more pronounced, while L357G results do not show a significant population of the pro-*S* conformer at large distances (Fig. S1B†). The correlation between the substrate spin density and its charge distribution was further verified by QM/MM calculations on 10 selected structures with different charge distribution (Fig. S2†).

In connection with the above computational results, improvement in the HRP sulfoxidation ability has been reported by removing the distal arginine residue (Fig. 1A).[33] A similar change is not possible in DyP engineering, since Arg332 (Fig. 1C) is necessary for DyP activation by H_2O_2, the catalytic

Fig. 6 Substrate positioning at the heme pocket: A) pro-*R* and pro-*S* MPS (left) and MTS (right) positioning on native DyP. B) Pro-*R* and pro-*S* MPS (left) and MTS (right) positioning on the F359G variant. Pro-*S* and pro-*R* substrate positions are shown as CPK sticks with the C atoms in green and cyan color, respectively. The spin density population on the substrate, from QM/MM calculations, is indicated in each image.
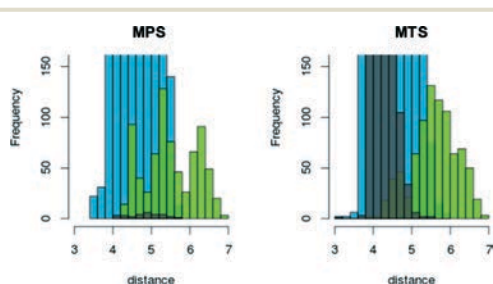

Fig. 7 Frequency–distance (in Å) distribution of MPS and MTS S atom to Arg332 CZ atom on the F359G variant for the selected reactive structures: last 10 kcal mol⁻¹ interaction energies and distances to the heme below 5 Å. Histograms for pro-*R* structures are shown in cyan, for pro-*S* structures in green, and for pro-*R/S* structures in grey.

efficiency being 2000-fold lower in the R332L variant, as shown above.

Conclusions

Stereoselective sulfoxidation ability could be introduced in a model (robust) peroxidase (*A. auricula-judae* DyP) by directed mutagenesis of residues at the distal side of the heme cofactor. The best variant (F359G) converted 95–99% MPS and MTS into the corresponding sulfoxides in 30 min reactions yielding the *S* enantiomer with 90–99 ee. The crystal structure of the F359G variant revealed an enlarged heme pocket enabling better accommodation of the sulfide substrates near the Fe^{4+}=O of the activated heme cofactor, whose oxygen atom is transferred to the sulfide substrate as shown by ^{18}O-labeling (in $H_2^{18}O_2$ reactions).

Computational modeling, based on the crystal structures of the parent enzyme and several variants, showed that the sulfoxidation ability of the F359G variant (no reaction was produced by the parent enzyme) is due to better substrate access (closer distance) to the reactive cofactor with better interaction energies, as shown by the PELE software. Moreover, the sulfoxidation stereoselectivity could be rationalized by QM/MM calculations predicting a higher reactivity of the sulfide molecules adopting pro-*S* poses at the active site (which was affected by a heme pocket arginine residue).

The study also demonstrates how molecular simulations can help obtain the required oxidation yields and stereoselectivities in sulfoxidation engineering.

Materials and methods

Chemicals

Hexane (HPLC quality), 4′-methoxyacetophenone, methylphenyl sulfoxide, *R*-(+)-methyl-*p*-tolyl sulfoxide, *R*/*S* methyl-*p*-tolyl sulfoxide, MPS and MTS were from Sigma-Aldrich. Isopropyl alcohol (LC-MS quality) was from Fluka. Dichloromethane (HPLC quality) was from Merck.

Directed mutagenesis

Simple DyP variants were produced by PCR using the pET23a-DyPI vector harboring the mature protein-coding sequence of *A. auricula-judae* DyP as a template.[24] For each mutation, a direct primer and a reverse primer were designed complementary to opposite strands of the DNA region containing the desired mutation. The sequences of the direct primers (with the mutated codons in italics) used for mutagenic PCR were the following: D168N mutation, 5′-G TTC GGC TTC CTT *AAC* GGA ATT GCT CAG CC-3′; R332L mutation, 5′-GCT AAC TCT ATC ATG *CTC* AGC GGC ATC CC-3′; L357G mutation, 5′-CT CAG GAG CGC GGC *GGA* GCG TTT GTG GCA TAC-3′; F359G, F359H or F359W mutations, 5′-GGC CTT GCG (*GGA*, *CAC* or *TGG*, respectively) GTG GCA TAC-3′.

PCR reactions were carried out in an Eppendorf (Hamburg, Germany) Mastercycler Pro using 10 ng of template DNA, 250 μM of each dNTP, 125 ng of direct and reverse

primers, 2.5 units of polymerase (Expand Long Template PCR System), and the manufacturer's reaction buffer. Reaction conditions were as follows: i) a "hot start" of 95 °C for 1 min; ii) 18 cycles at 95 °C for 50 s, 55 °C for 50 s, and 68 °C for 10 min; and iii) a final cycle at 68 °C for 10 min. The mutated sequences were confirmed by DNA sequencing using an ABI 3730 DNA analyzer (Applied Biosystem). pET23a-DyPI plasmids containing the mutations described above were digested with endonuclease DpnI and transformed into *E. coli* DH5α for propagation.

Enzyme production

E. coli BL21(DE3)pLysS cells with the pET23a-DyPI vector containing the *A. auricula-judae* mature DyP sequence (and those of DyP variants) were grown overnight at 37 °C and 170 rpm in Luria Bertani broth (with 100 μg mL^{-1} ampicillin and 34 μg mL^{-1} chloramphenicol) and used to inoculate 2 L flasks containing 1 L of Terrific Broth (TB) (with 100 μg mL^{-1} ampicillin and 34 μg mL^{-1} chloramphenicol) that were grown for 3 h at 37 °C and 200 rpm, induced with 1 mM isopropyl-β-D-thiogalactopyranoside, grown for further 4 h, and harvested by centrifugation. The apoenzyme, accumulated in inclusion bodies, was solubilized in 50 mM Tris-HCl (pH 8.0) containing 8 M urea, 1 mM EDTA, and 1 mM dithiothreitol for 1 h at 4 °C. *In vitro* activation of the native DyP and its D168N, R332L and L357G variants was carried out as previously described,[24] and specific activities were determined using 7.5 mM ABTS as a substrate and 2.5 mM H_2O_2.

For the three Phe359 variants, *in vitro* activation was performed at 4 °C using 0.15 M urea, 10 μM hemin, 0.02 mM dithiothreitol, 0.4 mM GSSG, 20% glycerol, 0.1 mM EDTA and 0.1 mg mL^{-1} protein in 50 mM phosphate (pH 6.5). After 144 h, the folding mixture was concentrated using a Millipore Pellicon ultrafiltration system (10 kDa cutoff) and centrifuged for 18 h at 13 000 rpm. Once concentrated, samples were dialyzed against 20 mM sodium acetate (pH 4.3), and the insoluble material was eliminated by centrifugation (13 000 rpm, 30 min).

Active DyPs (native enzyme and four directed variants) were purified using a Resource Q column (GE Healthcare) coupled to an ÄKTA liquid chromatography system, using a gradient from 0 to 0.3 M NaCl in 10 mM Tris-HCl, pH 7 (15 min, 2 mL min^{-1}). DyPs were analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis to confirm the purity of the proteins. The absorption spectra were recorded in 10 mM sodium tartrate (pH 5) at 25 °C using a Thermo Spectronic diode-array spectrophotometer. The DyP molar absorption coefficient (ε_{405} 117 000 M^{-1} cm^{-1}) was used for protein concentration determination.

Enzyme kinetics

Steady-state kinetic studies were performed by measuring the initial rates for MPS and MTS oxidation at 25 °C in 100 mM acetate, pH 5 (in triplicate). A 500 mM MPS and MTS stock in ethanol was serially diluted in ethanol to obtain final

6282 | *Catal. Sci. Technol.*, 2016, **6**, 6277–6285

This journal is © The Royal Society of Chemistry 2016

substrate concentrations from 250 mM to 2 mM. 10 μl of substrate were used to measure activity in 1 mL of reaction, in 1 cm light path cuvettes (when the initial absorbance was greater than 1, 0.33 mL reactions in 0.1 cm light path cuvettes were performed). 200 nM and 100 nM concentrations of native DyP and its mutated variants were used, respectively, and the reactions were started by adding 2 mM H_2O_2. The initial rates (absorbance decrease) were calculated using the differences in the molar absorption coefficients ($\Delta\varepsilon$) of MPS and methyl-phenyl sulfoxide ($\Delta\varepsilon_{254}$ 7870 M^{-1} cm^{-1}) and MTS and methyl-p-tolyl sulfoxide ($\Delta\varepsilon_{253}$ 8040 M^{-1} cm^{-1}).[41] Kinetic constants for H_2O_2 reduction and ABTS oxidation were estimated by following the formation of the ABTS cation radical (ε_{436} 29 300 M^{-1} cm^{-1}). In the former case, a saturation concentration of ABTS (7.5 mM for native DyP and the L357G variant and 2.5 mM for the D168N, R332L, F359G, F359H and F359W variants), 10 nM enzyme, and different concentrations of H_2O_2 (from 20 μM to 25 mM) in 100 mM tartrate, pH 3, were used, while in the second case oxidation of different ABTS concentrations, in the presence 2.5 mM H_2O_2, was measured. Plotting and analysis of kinetic curves were carried out using SigmaPlot 11.0. Apparent affinity, turnover number and catalytic efficiency were estimated by non-linear least-squares fitting to the Michaelis–Menten model.

Chromatographic analyses

Reactions (at 25 °C) were performed using 2 mL of 50 mM sodium acetate (pH 5) containing substrate (1 mM MPS or 0.5 mM MTS), 4 μM enzyme, and 1 mM H_2O_2. At 0, 30, 60 and 360 min of reaction, 0.5 mL aliquots were taken, and after addition of 4′-methoxyacetophenone as internal standard, the solution was extracted with hexane and analyzed by chiral HPLC.

The reaction products were analyzed using an Agilent HPLC equipment fitted with a Chiralpack IB column (dimensions of 4.6 mm × 250 mm and particle size of 5 μm) and as a mobile phase isocratic hexane : isopropyl alcohol (95 : 5, v/v) for 40 min for analysis of reactions with MPS and 98 : 2 (v/v) for 70 min for reactions with MTS, at a flow rate of 1 mL min^{-1}, at room temperature. Elution was monitored at 207, 216, 237, 248 and 263 nm. Calibration curves were obtained for identification and quantification of substrates and their sulfoxidation products. The retention times for the R (62 min) and S (64 min) methyl-p-tolyl sulfoxides were obtained from the corresponding standards, while those for the R (31 min) and S (33 min) methyl-phenyl sulfoxides were obtained from enzymatic oxidation of MPS assuming the published elution order,[42] which coincided with that observed for the two methyl-p-tolyl sulfoxides.

Enzymatic reactions with ^{18}O-labeled hydrogen peroxide ($H_2^{18}O_2$, 90% isotopic content, 2% w/v solution) and water ($H_2^{18}O$, 97% isotopic content) from Sigma-Aldrich were performed under the same conditions, extracted with dichloromethane, and analyzed by GC-MS using a gas chro-

matograph equipped with an HP-5MS column (Agilent, Santa Clara, CA, USA; 30 m × 0.25 mm internal diameter; 0.25 μm film thickness) coupled to a quadrupole mass detector. Helium was used as the carrier gas, and the oven program started at 110 °C for 2 min and increased at 20 °C min^{-1} until 240 °C. ^{18}O-labeling was calculated as the ratio between the abundances of the ^{18}O- to ^{16}O sulfoxide peaks, in their molecular ion (m/z 142 and 140, respectively, in the MPS reactions) chromatograms.

pH and temperature stability

The pH stability of the L357G and F359G variants and the native recombinant DyP was determined by incubating the enzymes (1 μM) for 24 h in 50 mM Britton–Robinson buffer (pH 2–12) at 25 °C. To evaluate their temperature stability, the enzymes in 10 mM sodium tartrate (pH 5) were incubated in the range from 25 to 80 °C for 10 min, followed by 2 min at 4 °C. The remaining activities were measured (in triplicate reactions) using 2.5 mM ABTS in 100 mM sodium tartrate (pH 3), as described above. The activity immediately after adding the enzyme to the buffer was taken as 100%. The T_{50} values, defined as the temperature at which 50% of activity is lost in 10 min incubation, were calculated.

Crystallization, data collection and refinement

Crystallization of the L357G and F359G variants was performed in 96-well sitting drop plates (Swissci MRC, England) at 22 °C using a Cartesian Honeybee robot (Digital, USA) and commercially available kits: JBScreen Classic (Jenna Bioscience, Germany), Wizard Classic Screen (Emerald Biostructures, USA) and ProPlex HT-96 (Molecular Dimensions, UK). Each droplet was 0.4 μL in size, containing 0.2 μL of protein (3 μg) solution and 0.2 μL of precipitant and was equilibrated over 50 μL of reservoir solution. Crystals of the L357G mutant were obtained in 5% MPD (v/v), 100 mM MES (pH 6.5) and 15% PEG 6000 (w/v). For the F359G variant, crystals were obtained in 2 M magnesium formate and 20% PEG 3350 (w/v). Crystals were cryoprotected using Paratone-N (Hampton Research).

X-ray diffraction images were collected at the ESRF (Grenoble, France) and ALBA (Barcelona, Spain) synchrotrons. Diffraction data were processed using XDS[43] and scaled using AIMLESS.[44,45] The structures were solved by molecular replacement using PHASER,[46] with the *A. auricula-judae* native DyP (PDB 4W7J) as the search model. The initial model was first refined with REFMAC5 (ref. 47) and alternating manual building with COOT.[48] The final model was obtained by repetitive cycles of refinement using PHENIX.[46] Subsequent refinement, introduction of solvent molecules and structure validation were as described for the native DyP[26] (data collection, refinement and final statistics of the two DyP variants are summarized in Table S1†).

The L357G structure did not show electron density for the first residue at the N-terminus, but the whole sequence could be solved for the F359G variant. In contrast, the C-terminal

This journal is © The Royal Society of Chemistry 2016

Catal. Sci. Technol., 2016, **6**, 6277–6285 | **6283**

region showed good electron density for both structures. The coordinates and structure factors have been deposited in the Protein Data Bank.

System preparation for molecular modeling

The starting structures for the computational simulations were the native DyP crystal at a resolution of 1.79 Å (PDB 4W7J)[26] and the crystal structures of the L357G and F359G variants (PDB 5IKG and 5IKD, respectively). Protein structures were prepared accordingly under pH 5 conditions, optimal for MPS and MTS sulfoxidation, using Schrödinger's Protein Preparation Wizard,[49] and the H++ web server.[50] Under these mild acidic conditions, histidines were double-protonated, with the exception of His-115 (ϵ-protonated) and His-304 (δ-protonated), and all other acidic residues were deprotonated. The heme site was modeled as thiolate-ligated compound I after being fully optimized in the protein environment with quantum mechanics/molecular mechanics (QM/MM) using QSite (see below for the level of theory).[51] Finally, MPS and MTS molecules were optimized with Jaguar,[52] at the density functional theory (DFT) M06 level with the 6-31G** basis set and Poisson–Boltzmann Finite element (PBF) implicit solvent in order to obtain their electrostatic potential atomic charges.

PELE computational analysis

Once the protein structures had been prepared and ligands optimized, heme binding site exploration was performed with PELE, a Monte Carlo-based algorithm capable of effectively sampling the protein–ligand conformational space.[53,54] The substrates were placed manually in identical positions at the entrance of the heme-access channel of each protein. PELE simulations were carried out in two stages: first, ligands were requested to move from the solvent to the heme site, and once the center of mass of the ligand was within 5 Å of the heme catalytic oxygen, it was free to explore the active site pocket with a 15 Å restrain. The results presented here are based on 240 trajectories × 48 h for each ligand.

QM/MM simulations

Hybrid QM/MM calculations were carried out in order to investigate the mutation effect on substrate sulfoxidation. PELE minima snapshots of MPS and MTS substrates on native DyP, L357G and F359F variant binding sites were selected for quantum calculations. For each case, at least two structures were investigated: a structure with the ligand in an optimal orientation toward the heme reactive oxygen to produce an R-(−)-sulfoxide product and a structure with the ligand placed correctly to give an S-(+)-sulfoxide product. QM/MM calculations were performed by including the heme (modeled as compound I), its axial ligand and the substrate in the quantum region and computing the spin density. In order to prepare the system for QM/MM, 0.5 ns molecular dynamics were performed with Desmond,[55] using SPC (sim-ple point charge) solvent and an ionic force of 0.15 M. Calculations were performed at the DFT M06-L(lacvp*)/OPLS level with QSite.[52]

Acknowledgements

Notes and references

1 I. Fernández and N. Khiar, *Chem. Rev.*, 2003, **103**, 3651–3706.

2 H. Gröger, in *Catalytic Asymmetric Synthesis*, ed. I. Ojima, John Wiley & Sons, Inc., 2010, pp. 269–341.

3 W. J. H. van Berkel, N. M. Kamerbeek and M. W. Fraaije, *J. Biotechnol.*, 2006, **124**, 670–689.

4 S. Colonna, N. Gaggero, C. Richelmi and P. Pasta, *Trends Biotechnol.*, 1999, **17**, 163–168.

5 A. Matura and K. H. van Pee, in *Enzyme catalysis in organic synthesis: A comprehensive handbook*, ed. K. Drautz, H. Gröger and O. May, Wiley, Weinheim, 2012, pp. 1553–1567.

6 E. G. Hrycay and S. M. Bandiera, *Arch. Biochem. Biophys.*, 2012, **522**, 71–89.

7 T. Matsui, Y. Dekishima and M. Ueda, *Appl. Microbiol. Biotechnol.*, 2014, **98**, 7699–7706.

8 S. Colonna, N. Gaggero, L. Casella, G. Carrea and P. Pasta, *Tetrahedron: Asymmetry*, 1992, **3**, 95–106.

9 M. P. J. van Deurzen, F. van Rantwijk and R. A. Sheldon, *Tetrahedron*, 1997, **53**, 13183–13220.

10 S. Kobayashi, M. Nakano, T. Goto, T. Kimura and A. P. Schaap, *Biochem. Biophys. Res. Commun.*, 1986, **135**, 166–171.

11 S. Colonna, N. Gaggero, G. Carrea and P. Pasta, *J. Chem. Soc., Chem. Commun.*, 1992, 357–358.

12 R. Z. Harris, S. L. Newmyer and P. R. Ortiz de Montellano, *J. Biolumin. Chemilumin.*, 1993, **268**, 1637–1645.

13 V. M. Dembitsky, *Tetrahedron*, 2003, **59**, 4701–4720.

14 M. Sono, M. P. Roach, E. D. Coulter and J. H. Dawson, *Chem. Rev.*, 1996, **96**, 2841–2888.

15 V. P. Miller, G. D. DePillis, J. C. Ferrer, A. G. Mauk and P. R. Ortiz de Montellano, *J. Biolumin. Chemilumin.*, 1992, **267**, 8936–8942.

16 W. Adam, C. Mock-Knoblauch and C. R. Saha-Möller, *J. Organomet. Chem.*, 1999, **64**, 4834–4839.

17 E. Baciocchi, M. F. Gerini, P. J. Harvey, O. Lanzalunga and S. Mancinelli, *Eur. J. Biochem.*, 2000, **267**, 2705–2710.

18 M. Hofrichter, H. Kellner, M. J. Pecyna and R. Ullrich, *Adv. Exp. Med. Biol.*, 2015, **851**, 341–368.

6284 | *Catal. Sci. Technol.*, 2016, **6**, 6277–6285

This journal is © The Royal Society of Chemistry 2016

19 A. Horn, R. Ullrich, M. Hofrichter and K. Scheibner, *Abstracts BioTrans-2007*, Oviedo, 2007, p. 63.

20 A. Horn, *Der Einsatz einer neuartigen Peroxidase aus dem Basidiomyceten Agrocybe aegerita am Beispiel der enantioselektiven Sulfoxidation*, PhD thesis, Universität Rostock, Rostock, 2009.

21 E. Aranda, M. Kinne, M. Kluge, R. Ullrich and M. Hofrichter, *Appl. Microbiol. Biotechnol.*, 2009, **82**, 1057–1066.

22 B. Goblirsch, R. C. Kurker, B. R. Streit, C. M. Wilmot and J. L. Dubois, *J. Mol. Biol.*, 2011, **408**, 379–398.

23 M. Puhse, R. T. Szweda, Y. Y. Ma, C. Jeworrek, R. Winter and H. Zorn, *Biochim. Biophys. Acta, Proteins Proteomics*, 2009, **1794**, 1091–1098.

24 D. Linde, C. Coscolín, C. Liers, M. Hofrichter, A. T. Martínez and F. J. Ruiz-Dueñas, *Protein Expression Purif.*, 2014, **103**, 28–37.

25 D. Linde, F. J. Ruiz-Dueñas, E. Fernández-Fueyo, V. Guallar, K. E. Hammel, R. Pogni and A. T. Martínez, *Arch. Biochem. Biophys.*, 2015, **574**, 66–74.

26 D. Linde, R. Pogni, M. Cañellas, F. Lucas, V. Guallar, M. C. Baratto, A. Sinicropi, V. Sáez-Jiménez, C. Coscolín, A. Romero, F. J. Medrano, F. J. Ruiz-Dueñas and A. T. Martínez, *Biochem. J.*, 2015, **466**, 253–262.

27 E. Strittmatter, C. Liers, R. Ullrich, S. Wachter, M. Hofrichter, D. A. Plattner and K. Piontek, *J. Biolumin. Chemilumin.*, 2013, **288**, 4095–4102.

28 C. Liers, C. Bobeth, M. Pecyna, R. Ullrich and M. Hofrichter, *Appl. Microbiol. Biotechnol.*, 2010, **85**, 1869–1879.

29 J. N. Rodríguez-López, A. T. Smith and R. N. F. Thorneley, *J. Biolumin. Chemilumin.*, 1996, **271**, 4023–4030.

30 K. Piontek, E. Strittmatter, R. Ullrich, G. Grobe, M. J. Pecyna, M. Kluge, K. Scheibner, M. Hofrichter and D. A. Plattner, *J. Biolumin. Chemilumin.*, 2013, **288**, 34767–34776.

31 S. L. Newmyer and P. R. Ortiz de Montellano, *J. Biolumin. Chemilumin.*, 1995, **270**, 19430–19438.

32 S. Ozaki and P. R. Ortiz de Montellano, *J. Am. Chem. Soc.*, 1995, **117**, 7056–7064.

33 M. I. Savenkova, J. M. Kuo and P. R. Ortiz de Montellano, *Biochemistry*, 1998, **37**, 10828–10836.

34 M. I. Savenkova and P. R. Ortiz de Montellano, *Arch. Biochem. Biophys.*, 1998, **351**, 286–293.

35 F. van Rantwijk and R. A. Sheldon, *Curr. Opin. Biotechnol.*, 2000, **11**, 554–564.

36 E. van Bloois, D. E. T. Pazmino, R. T. Winter and M. W. Fraaije, *Appl. Microbiol. Biotechnol.*, 2010, **86**, 1419–1430.

37 Y. K. Bong, M. D. Clay, S. J. Collier, B. Mijts, M. Vogel, X. Zhang, Y. Zhu, J. Nazor, D. Smith and S. Song, *Patent (International)*, WO071982, 2011.

38 S. Shaik, Y. Wang, H. Chen, J. Song and R. Meir, *Faraday Discuss.*, 2010, **145**, 49–70.

39 Y. Goto, T. Matsui, S. I. Ozaki, Y. Watanabe and S. Fukuzumi, *J. Am. Chem. Soc.*, 1999, **121**, 9497–9502.

40 P. Molina-Espeja, M. Cañellas, F. Plou, M. Hofrichter, F. Lucas, V. Guallar and M. Alcalde, *ChemBioChem*, 2016, **17**, 341–349.

41 C. Capeillere-Blandin, C. Martin, N. Gaggero, P. Pasta, G. Carrea and S. Colonna, *Biochem. J.*, 1998, **335**(Pt 1), 27–33.

42 Y. Wang, H. Li, W. Qi, Y. Yang, Y. Yan, B. Li and L. Wu, *J. Mater. Chem.*, 2012, **22**, 9181–9188.

43 W. Kabsch, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 125–132.

44 P. R. Evans, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2011, **67**, 282–292.

45 P. R. Evans and G. N. Murshudov, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2013, **69**, 1204–1214.

46 A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658–674.

47 G. N. Murshudov, A. A. Vagin and D. J. Dodson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1997, **53**, 255.

48 P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 486–501.

49 G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, *J. Comput.-Aided Mol. Des.*, 2013, **27**, 221–234.

50 R. Anandakrishnan, B. Aguilar and A. V. Onufriev, *Nucleic Acids Res.*, 2012, **40**, W537–W541.

51 Schrödinger, *QSite 5.7*, LCC, New York, 2011.

52 Schrödinger, *Jaguar 8.1*, LCC, New York, 2013.

53 K. W. Borrelli, A. Vitalis, R. Alcantara and V. Guallar, *J. Chem. Theory Comput.*, 2005, **1**, 1304–1311.

54 K. Borrelli, B. Cossins and V. Guallar, *J. Comput. Chem.*, 2010, **31**, 1224–1235.

55 Desmond, *Maestro-Desmond Interoperability Tools, version 2.4*, Schrödinger, New York, NY, 2010.

This journal is © The Royal Society of Chemistry 2016

Catal. Sci. Technol., 2016, **6**, 6277–6285 | **6285**

-140-

Supporting Information

Asymmetric sulfoxidation by engineering the heme pocket of a dye-decolorizing peroxidase: An experimental and computational study

Dolores Linde, Marina Cañellas, Cristina Coscolín, Irene Davó-Siguero, Antonio Romero, Fátima Lucas, Francisco J. Ruiz-Dueñas, Victor Guallar and Angel T. Martínez

Table S1 X-ray data collection and refinement statistics of L357G and F359G variants of *A. auricula-judae* DyP

	L357G	F359G
Data collection		
Beamline	ID23-1 (ESRF)	BL13-XALOC (ALBA)
Wavelength (Å)	0.9792	0.9795
Space group	C2	C2
Unit cell parameters (Å,°)	a=100.9, b=61.0, c=81.0; β=95.0	a=104.9, b=56.1, c=82.7; β=97.03
Resolution range (Å)	44.73 - 1.95 (2.05 - 1.95)	46.63 - 1.10 (1.17 - 1.10)
Total reflections	239,457 (35,489)	997,210 (135,507)
Unique reflections	35,827 (5,247)	184,770 (26,363)
Multiplicity	6.7 (6.8)	5.4 (5.1)
Completeness (%)	99.2 (99.9)	98.4 (96.4)
Mean $I/\sigma(I)$	12.6 (2.3)	9.5 (2.4)
R_{merge}[a]	0.073 (0.729)	0.088 (0.726)
$CC_{1/2}$ (%)[b]	99.8 (90.2)	99.7 (77.2)
Mosaicity (°)	0.16	0.15
Wilson B factor (Å^2)	33.21	7.78
Refinement		
R_{work}/R_{free}	0.198 (0.367) / 0.23 (0.415)	0.149 (0.231) / 0.158 (0.246)
Working reflections	33,925 (5,206)	182,839 (25,500)
Testing reflections	1,863 (308)	1,897 (280)
Protein atoms (non H)	3,298	3,310
Heme group	43	43
Water molecules	127	576
Mean B factors (Å^2):		
-Protein atoms (non H)	46.2	11.9
-Heme group	41.7	7.8
-Water molecules	47.5	24.7
Deviations from ideality:		
-rmsd bond lengths (Å)	0.007	0.006
-rmsd angles (°)	0.876	0.987
Ramachandran statistics:		
-Favoured (%)	97.5	98.9
-Outliers (%)	0	0
PDB code	5IKG	5IKD

Statistics for the highest-resolution shell are shown in parenthesis. [a]$R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - <I(hkl)>| / \Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity measured for the *i*th reflection and *(I(hkl))* is the average intensity of all reflections with indices *hkl*. [b]$CC_{1/2}$ is the correlation coefficient between two random half datasets (Karplus P.A and Diederichs K., Science, 2012, 336, 1030-1033).
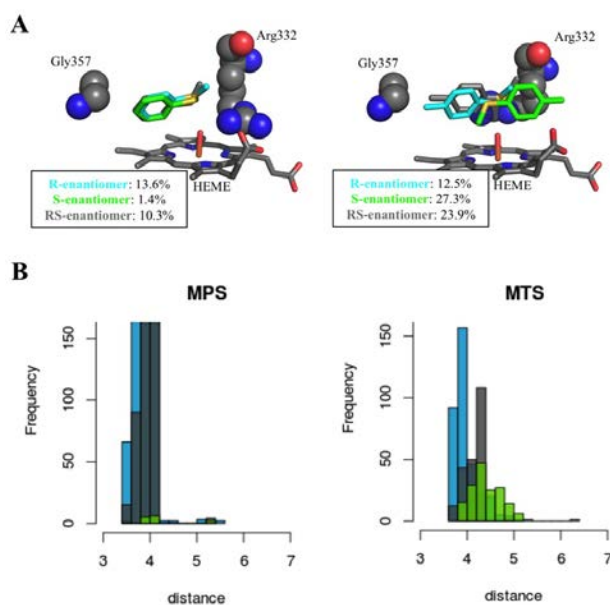
Fig. S1. L357G variant substrate oxidation and positioning analysis: **A)** pro-*R*, pro-*S* and pro-*RS* MPS (left) and MTS (right) positioning on the L357G variant. Spin density population on the substrate, from QM/MM calculations, is indicated in each image. **B)** Frequency distance (in Å) distribution of MPS and MTS S atom to Arg332 CZ atom in the L357G variant for the selected reactive structures: last 10 kcal·mol⁻¹ interaction energies and distances to the heme group below 5 Å. Pro-*S* substrate positions are colored in green, pro-*R* in cyan and pro-*RS* in gray.
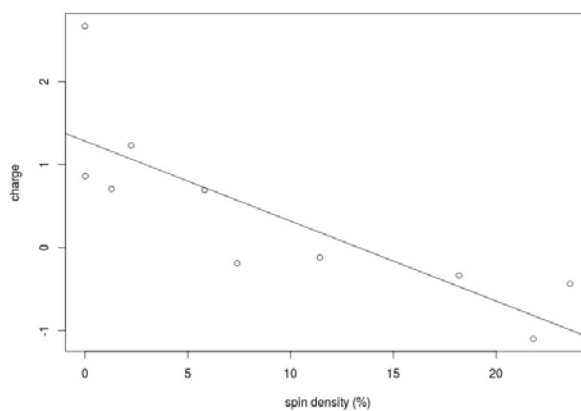

Fig. S2. Correlation between the substrate spin density and the charge distribution on 10 representative points.

3.2.2 *Article V:* Synthesis of 1-naphthol by a natural peroxygenase engineered by directed evolution.

Patricia Molina-Espeja,[a] Marina Cañellas,[b] Francisco J. Plou,[a] Martin Hofrichter,[c] Fatima Lucas,[b, d], Victor Guallar,[d, e] and Miguel Alcalde*[a]

[a] Deparment of Biocatalysis, Institute of Catalysis, CSIC, C/Marie Curie nº2L10, 28049 Madrid (Spain) E-mail: **malcalde@icp.csic.es**

[b] Anaxomics Biotech, Balmes 89, 08008 Barcelona (Spain)

[c] TU Dresden, International Institute Zittau, Department of Bio- and Environmental Sciences, Markt 23, 02763 Zittau (Germany)

[d] Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, 08034 Barcelona (Spain)

[d] ICREA: Institució Catalana de Recerca i Estudis Avançats, Passeig Lluís Companys 23, 08010 Barcelona (Spain)

Synthesis of 1-Naphthol by a Natural Peroxygenase Engineered by Directed Evolution

Patricia Molina-Espeja,[a] Marina Cañellas,[b] Francisco J. Plou,[a]Martin Hofrichter,[c] Fatima Lucas,[b, d] Victor Guallar,[d, e] and Miguel Alcalde*[a]

[a]   P. Molina-Espeja, Prof. F. J. Plou, Prof. M. Alcalde Department of Bio-catalysis, Institute of Catalysis, CSIC C/Marie Curie n8 2L10, 28049 Madrid (Spain)
E-mail: malcalde@icp.csic.es

[b]   M. Cañellas, Dr. F. Lucas Anaxomics Biotech, Balmes 89, 08008 Barcelona (Spain)

[c]   Prof. M. Hofrichter. TU Dresden, International Institute Zittau Department of Bio- and Environmental Sciences Markt 23, 02763 Zittau (Germany)

[d]   Dr. F. Lucas, Prof. V. Guallar. Joint BSC-CRG-IRB Research Program in Computational Biology Barcelona Supercomputing Center, Jordi Girona 29, 08034 Barcelona (Spain)

[e]   Prof. V. Guallar, ICREA: Instituci6 Catalana de Recerca i Estudis Avançats Passeig Lluís Companys 23, 08010 Barcelona (Spain)

There is an increasing interest in enzymes that catalyze the hydroxylation of naphthalene under mild conditions and with minimal requirements. To address this challenge, an extracellular fungal aromatic peroxygenase with mono(per)oxygenase activity was engineered to convert naphthalene selectively into 1-naphthol. Mutant libraries constructed by random mutagenesis and DNA recombination were screened for peroxygenase activity on naphthalene together with quenching of the undesired peroxidative activity on 1-naphthol (one-electron oxidation). The resulting double mutant (G241D-R257K) obtained from this process was characterized biochemically and computationally. The conformational changes produced by directed evolution improved the substrate's catalytic position. Powered exclusively by catalytic concentrations of H_2O_2, this soluble and stable biocatalyst has a total turnover number of 50 000, with high regioselectivity (97 %) and reduced peroxidative activity.

Introduction

Currently, around 40 000 tons of 1-naphthol (naphthalen-1-ol) are generated each year to fulfill the need to produce herbicides, insecticides, pharmaceuticals, and dye precursors.[1,2] The manufacturing process involves the use of hazardous chemical catalysts with low turnover numbers and poor regioselectivity, and it is associated with elevated energy consumption, high costs, and the release of harmful waste products.[3–7] Oxygenases that perform regioselective oxyfunctionalization of aromatic rings would offer a greener alternative to production by standard chemical methods. Accordingly, the manufacture of 1- naphthol with the aid of enzymes should ideally take place in a one-pot process under mild conditions [room temperature, atmospheric pressure, and essentially aqueous solution (with small amounts of organic co-solvents)], while reducing both the energy required and the noxious waste products that result from chemical synthesis.[8–11] To date, most studies of the biocatalytic synthesis of 1-naphthol have focused on monooxygenases. In particular, over the years P450 monooxygenases have been designed for different purposes, from the selective hydroxylation of alkanes (including terminal hydroxylation) to the unnatural olefin cyclopropanation by carbene transfer.[12–15] These enzymes can transform naphthalene into 1-naphthol either by harnessing the peroxide shunt pathway or through their natural NAD(P)H-dependent activity. More recently, directed evolution of toluene *ortho*-monooxygenase (TOM) with a whole-cell biocatalytic system was described.[16–18] However, poor enzyme stability and reliance on expensive redox cofactors and reductase domains have so far precluded the practical application of these enzymes in specific industrial settings.

Over a decade ago, the first "true natural" aromatic peroxygenase was discovered (EC 1.11.2.1; also referred to as unspecific peroxygenase, UPO).[19] This enzyme was recently classified as a member of the new heme-thiolate peroxidase superfamily (HTP), along with chloroperoxidase (CPO) from *Caldariomyces fumago*, even though CPO is not capable of transferring oxygen to aromatic rings or *n*-alkanes.[20] The properties of UPO resemble those of P450s in terms of the selective C−H oxyfunctionalization of organic compounds. However, UPO is an extracellular, highly active, and stable enzyme, and it does not depend on expensive redox cofactors or auxiliary flavoproteins. UPO is "fueled" by stoichiometric concentrations of H_2O_2, which serves both as an enzyme co-oxidant (i.e., primary electron acceptor) and as a source of oxygen. Thus, with minimal requirements UPO is capable of performing diverse transformations of great complexity and relevance for organic synthesis, such as the hydroxylation of aromatic and aliphatic com- pounds, the epoxidation of olefins, the N- and S-oxidation of heterocycles, or the cleavage of ethers, to name but a few.[21] Over 300 substrates (a number that is still growing)

for UPO have already been reported; among these, naphthalene can be converted into 1-naphthol through an epoxide intermediate (as previously described for P450s). Apart from its natural mono(per)oxygenase activity through two-electron monooxygenation, similar to that in the P450 shunt pathway (confirmed by experiments with ^{18}O-labeled H_2O_2), UPO also displays peroxidative activity towards phenolic substrates and other com- pounds through a one-electron oxidation route.[22] This convergence of two such activities in the same protein could become a problem from the point of view of industrial applications, because UPO hydroxylation products are always found along with varying amounts of oxidation byproducts. This is especially true for aromatic hydroxylation reactions in which the product(s) (phenolics) released by peroxygenase activity can serve as reducing substrate(s) of the peroxidative activity, the latter promoting the formation of phenoxyl radicals and quinones that can undergo undesired nonenzymatic polymerization and affect the overall production yields.

In this work we have employed directed evolution to tailor a UPO from the edible mushroom *Agrocybe aegerita* (*Aae*UPO1) for the efficient regioselective synthesis of 1-naphthol. Mutant libraries were constructed and expressed in yeast, yielding a final variant that was characterized comprehensively, and its use in the reaction to synthesize 1-naphthol was studied in depth. To obtain a detailed atomic explanation of the effects of the mutations, we also performed ligand migration and quantum chemical simulations.

Results and Discussion

Directed evolution approach

Our departure point was a UPO mutant (PaDa-I) that we had previously created by directed evolution in *Saccharomyces cerevisiae*.[23] PaDa-I harbors nine mutations (F12Y-A14V-R15G- A21D-V57A-L67F-V75I-I248V-F311L: the mutations in the signal peptide are underlined) that enhance its functional expression in yeast (8 mg L^{-1} in *S. cerevisiae* and over 200 mg L^{-1} in *Pichia pastoris*).[24] At the same time this mutant retains strong activity and stability, particularly in terms of temperature and the presence of co-solvents. Here, mutant libraries of UPO were con- structed by random mutagenesis, staggered extension process (StEP) recombination, and in vivo shuffling and were explored for 1-naphthol synthesis. A sensitive ad hoc dual assay was designed to cap the peroxidative activity on 1-naphthol while protecting the peroxygenase activity on naphthalene (**Figure 1**).

After only two rounds of directed evolution (≈ 4000 clones screened), we identified the double mutant G241D-R257K (termed JaWa), which displayed a twofold improvement in peroxygenase activity and half the peroxidative activity of the parent (**Figure S1** in the Supporting Information).
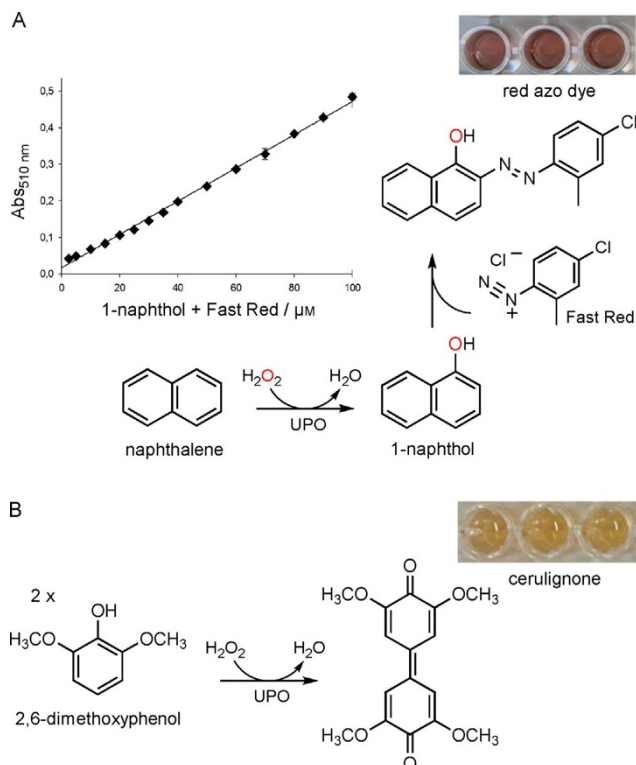

Figure 1. Dual screening assay for naphthalene transformation by UPO. A) For positive selection (peroxygenase activity), mutant libraries were screened with naphthalene such that the formation of the colorless 1-naphthol could be detected by use of the Fast Red reagent.[25] This compound couples specifically to 1-naphthol to form a red azo dye that can be readily measured at 510 nm ($\varepsilon_{510}= 4700 \, M^{-1}cm^{-1}$), at which interference with the culture broth is minimal (the inset shows the linearity and the limit of detection of the assay). B) For the negative criterion of the assay (peroxidative activity), we used a surrogate substrate (the aromatic 2,6-dimethoxyphenol, DMP) that can be oxidized by UPO to form the orange-colored cerulignone ($\varepsilon_{469}=27500 \, M^{-1}cm^{-1}$). The coefficient of variance for both assays was below 12%.

Biochemical characterization

PaDa-I and JaWa variants were produced, purified to homogeneity [Reinheitszahl, $Rz(A_{418}/A_{280}) \approx 2$] and characterized biochemically. The general spectral characteristics, molecular weight, N-terminal processing, and degree of glycosylation were maintained for both the parent and the mutant enzymes (**Table 1**).

Table 1. Biochemical features of PaDa-I and JaWa variants.		
Biochemical and spectroscopy features	PaDa-I	JaWa
MW [Da][a]	52 000	52 000
MW [Da][b]	51 100	51 100
MW [Da][c]	35 914	35 944
degree of glycosylation [%]	30	30
N-terminal end	EPGLPPPGPL	EPGLPPPGPL
thermal stability, T_{50} [8C][d] pl	57.6	59.7
optimum pH for ABTS optimum	5.5	5.3
pH for DMP optimum pH for	4.0	4.0
naphthalene Rz (A_{418}/A_{280})	6.0	6.0
Soret region [nm] CT1	6.0	6.0
[nm]	1.8	2.3
CT2 [nm]	418	418
	570	570
	537	537

[a] Estimated by SDS-PAGE; [b] estimated by MALDI-TOF mass spectrome- try; [c] estimated from amino acid composition; [d] estimated from cul- ture superna-tants.

JaWa showed improved thermostability (with a 28 °C increase in the T_{50} value, defined as the temperature at which the enzyme retains 50 % of its activity after a 10 min incubation) and strong stability in the presence of the acetonitrile that is used to keep naphthalene in solution (solubility in pure water 31.7 mg L^{-1}, **Figure S2**). Naphthalene transformation by JaWa and by PaDa-I was analyzed by HPLC-PDA (photodiode array detector). Naphthalene oxygenation by *Aae*UPO1 occurs through an unstable naphthalene 1,2-oxide intermediate, which undergoes rapid hydrolysis at acid pH into naphthol (naphthalen-1- and -2-ol).[22] Accordingly, we first measured the product distribution after stopping a 15 min reaction with HCl. Both PaDa-I and JaWa showed similar regioselectivity (92 % 1- naphthol, 8 % 2-naphthol), whereas JaWa produced significantly more 1-naphthol (156 % relative to PaDa-I) without any formation of the 1-naphthol oxidation product—1,4-naph-thoquinone—being detectable (**Figure 2 A**). Similar behavior was observed when we followed the reaction for longer (270 min at pH 7.0 without stopping the reac-tion by adding HCl); this indicated that the chemical conversion of naphthalene 1,2-oxide into naphthols also occurs at neutral pH, although at lower rates. More-over, traces of 1,4-naphthoquinone were detected (**Figure 2 B–D**). Whereas the formation of the epoxide inter- mediate by both enzymes reached a plateau at ≈ 40 min (due to oxidative damage caused by H$_2$O$_2$), the regioselectivity was enhanced to 97 % 1-naphthol. This result is in good agreement with the loss of selectivity ob-served when HCl is used to stop the reaction, due to the increased reactivity of the epoxide intermediate in an acid environment [22]. In terms of production yields, the differences between the mutants were even more pronounced (i.e., 0.14 and 0.32 mm 1-naphthol for PaDa-I and JaWa, respectively), whereas the products

produced were similar when analyzed by mass spectrometry (**Figure S3**). We cal-culated total turnover numbers (TTNs) of approximately 50 000 for JaWa as op-posed to 20 000 for the PaDa-I parent. When the kinetic parameters of the peroxy-genative and peroxidative activities of both enzymes were determined (**Table 2**), the k_{cat}/K_m value (catalytic efficiency) of JaWa for naphthalene hydroxylation was increased 1.5-fold, whereas the peroxidative activity was partially suppressed, with dramatic three- to 12-fold decreases in catalytic efficiencies for the peroxidative substrates 2,6-DMP and ABTS [2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)], respectively. Moreover, the k_{cat}/K_m value of H_2O_2 with benzyl alcohol as reducing substrate was also affected.
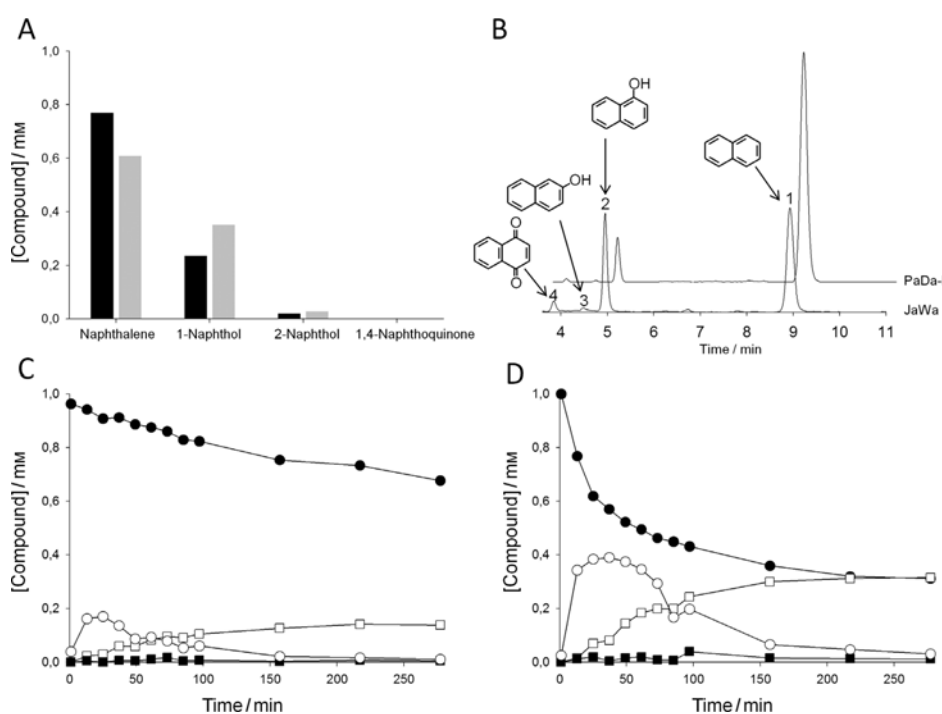

Figure 2. Naphthalene conversion by mutant UPOs. A) Product distribution after a 15 min reaction stopped with HCl (PaDa-I, black bars; JaWa, gray bars). Reactions were performed at room temperature, and each reaction mixture contained 6.6 nm purified enzyme, 1 mm naphthalene, 20 % acetonitrile, and 1 mm H2O2 in 100 mm potassium phosphate buffer pH 7.0 (1 mL final volume). After 15 min, the reaction was stopped with 20 mL of 37 % HCl and the mixture was analyzed by HPLC-PDA. B) HPLC elution profiles after a reac-tion time of 270 min: 1) naphthalene, 2) 1-naphthol, 3) 2-naphthol, 4) 1,4-naphthoquinone. C) and D) Time courses of the reactions over 270 min at pH 7.0 with PaDa-I and JaWa, respectively (without stopping of the reactions by addition of HCl). ●: naphthalene, ○: naph-thalene 1,2-oxide, ⊡: 1-naphthol, ■: 2-naphthol. The total turnover numbers (TTNs,

reported as mmol product/mmol enzyme) were estimated from the 1-naphthol concentration after 270 min.
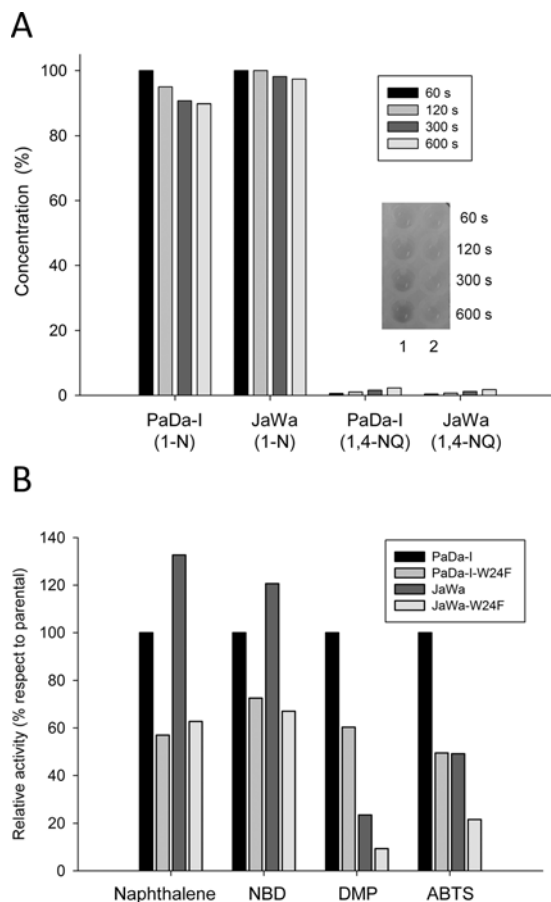

Figure 3. A) Turnover rates for 1-naphthol oxidation. Reactions (1 mL final volume) were carried out at room temperature, and each mixture contained 1 mM naphthol, 1 mm H_2O_2, 20% acetonitrile, and purified enzyme (40 nM) in 100 mM potassium phosphate pH 7.0. 1-N: 1-naphthol. 1,4-NQ: 1-4-naphthoquinone. The reactions were performed in triplicate and were stopped by addition of HCl (pH<1) after different times (60 to 600 s). Inset: colored polymeric products from 1,4-naphthoquinone. 1: PaDa-I. 2: JaWa. B) Activities of W24F site-directed variants. The buffer used was 100 mM potassium phosphate pH 7.0, except for ABTS, for which the buffer was 100 mM sodium phosphate/citrate pH 4.0. The substrate concentrations were: 0.5 mM naphthalene, 1 mM NBD, 3 mM DMP, and 0.3 mM ABTS. In all cases, 1 mM H_2O_2 and 15% acetonitrile were added to complete the reaction mixture. For naphthalene activity, a Fast Red assay was used (after 10 min of reaction, Fast Red was added at a final concentration of 0.5 mM, and when a red color appeared and was stabilized, the absorbance was measured). The molar extinction coefficients are: ε_{510}=4700 $M^{-1}cm^{-1}$ for naphthalene+Fast Red, ε_{425}=9700 $M^{-1}cm^{-1}$ for NBD, ε_{469}=27500 $M^{-1}cm^{-1}$ for DMP, and ε_{418}=36000 $M^{-1}cm^{-1}$ for ABTS.

To ascertain how the decrease in the peroxidative activity affected the TTN for the hydroxylation of naphthalene, we measured the turnover rate [μmol product per μmol enzyme per min] for the conversion of 1-naphthol into 1,4-naphthoquinone by HPLC. Although the catalytic rate of PaDa-I for 1-naphthol was already low (200 min^{-1}), it dropped further to 92 min^{-1} for JaWa, with a $\approx$ 1.5-fold reduction in the 1,4-naphthoquinone/ 1-naphthol molar ratio (Figure 3 A). This effect could also be visualized, because the polymeric products formed through the non-enzymatic coupling of the phenoxyl radicals produced by PaDa-I are colored (inset in **Figure 3A**).[12]

Table 2. Kinetic constants for the PaDa-I and JaWa variants.			
Substrate	Kinetic constants	PaDa-I	JaWa
ABTS	K_m [mm]	48.0 ± 4.5	181 ± 22
	k_{cat} [s^{-1}]	395 ± 13	125 ± 5
	k_{cat}/K_m [s^{-1}m^{-1}]	8.2 x 10^6± 6x10^5	6.9x 10^5±6.3x10^4
DMP	K_m [mm]	126 ± 14	866 ± 108
	k_{cat} [s^{-1}]	68 ± 2	142 ± 8
	k_{cat}/K_m [s^{-1}m^{-1}]	5.4x 10^5±4.8x10^4	1.6x 10^5±1.2x10^4
naphthalene	K_m [mm]	578 ± 106	127 ± 27
	k_{cat} [s^{-1}]	229 ± 17	78 ± 3
	k_{cat}/K_m [s^{-1}m^{-1}]	4.0x 10^5±4.0x10^4	6.2x 10^5±1.1x10^5
H$_2$O$_2$	K_m [mm]	486 ± 55	1250 ± 300
	k_{cat} [s^{-1}]	238 ± 8	447 ± 40
	k_{cat}/K_m [s^{-1}m^{-1}]	5.0x 10^5±4.2x10^4	3.6x 10^5±5.9x10^4

The kinetic constants for ABTS were measured in 100 mm sodium phos- phate/citrate buffer pH 4.0 containing 2 mm H$_2$O$_2$, whereas the rest of the substrates were tested in 100 mm potassium phosphate pH 7.0 contain- ing 2 mm H$_2$O$_2$ (DMP) or 1 mm H$_2$O$_2$. Varying amounts of naphthalene were added from stocks in acetoni- trile to give a final co-solvent concen- tration of 20%, v/v. For H$_2$O$_2$, benzyl alco- hol was used as a reducing sub- strate under the corresponding saturated condi- tions.

It has been reported that UPO might resemble CPO in terms of the presence of different oxidation sites for peroxidative activity.[26] Thus, we examined the recent *Aae*UPO1 crystal structure to find a means to suppress alternative peroxidative pathways.[27] QM/MM calculations identified Trp24 as the most oxidizable surface residue (see computational analysis below); hence, we constructed site-directed variants by using the PaDa-I and JaWa variants as templates. The corresponding PaDa-I-W24F and JaWa-W24F mutants were compared: irrespective of the variant and the reducing substrate tested, the W24F mutation decreased the peroxidative activity by $\approx$ 60%(**Figure 3B**). However, this mutation also affected the peroxygenase activity, with a $\approx$ 50 % drop in the hydroxylation of naphthalene and the fission of 5-nitro-1,3-benzodioxole (NBD), thus suggesting that Trp24 might also play a role in the peroxygenase activity of UPO.

Computational analysis

The JaWa mutations were mapped in the *Aae*UPO1 structure with a characteristic binding pocket in which the hydrophobic Phe69-Phe121-Phe199 triad is involved in the correct orienta- tion of aromatic compounds (**Figures 4 and S4**). The G241D mutation is located at the entrance of the heme access chan- nel, whereas R257K is situated at the surface of the protein, far from any relevant catalytic sites. In order to improve our un- derstanding of the atomic mechanistic details responsible for the observed differences between PaDa-I and JaWa variants, we turned to molecular modeling, using the PELE (protein energy landscape exploration) software[28] and QM/MM methods[29] (see the Supporting Movie).
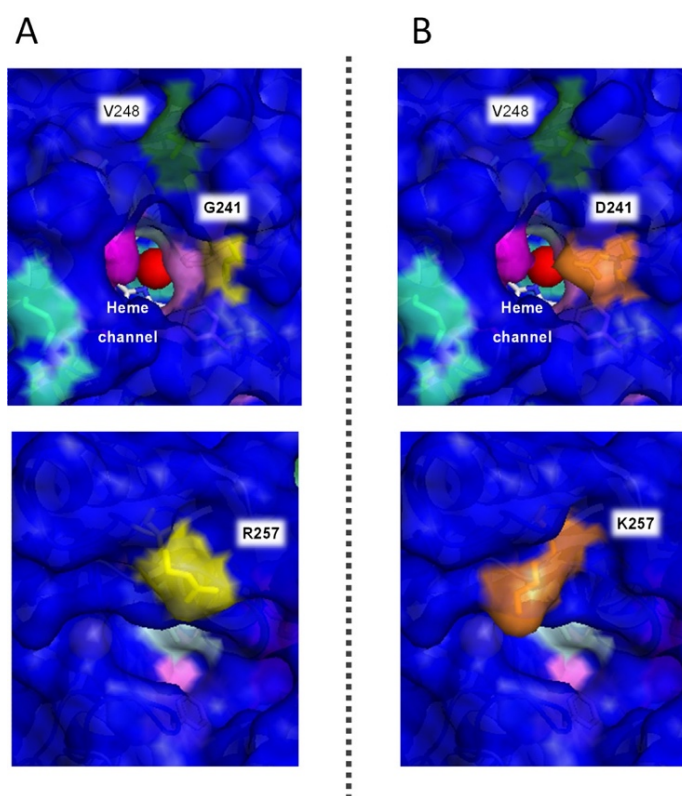
Figure 4. Mutations in the evolved UPO. Protein modeling was based on the AaeUPO1 crystal structure (PDB accession number 2YOR). A) PaDa-I. B) JaWa. The G241D and R257K substitutions are depicted in yellow and orange. The V248 mutation (dark green) comes from previous directed evolution investigations.[23] The Phe residues involved in accommodating substrates at the catalytic pocket are highlighted in pink, the Cys residues are in light green (C36 axial ligand and the C278–C319 disulfide bridge), the R189 com- ponent of the acid-base pair involved in the catalysis is tagged in light blue, and the Fe^{3+} of heme is represented as a red sphere.

Peroxidative activity: Global ligand migration exploration with PELE, aimed at identifying the most favorable binding sites, found the heme entrance, Trp24, and Tyr29 as the most favorable minima. Then, by the protocol previously employed to identify catalytically active surface residues in a dye-decolorizing peroxidase,[30] QM/MM calculations were performed on wild-type *Aae*UPO1. All potential surface oxidation sites (nine tyrosine residues and one tryptophan) were included in the quantum region, with one electron being removed and spin density computed. Furthermore, pairwise comparison was performed with the residues that showed a clear preference to be oxidized: Trp24 and Tyr47. Results showed Trp24 to be the most favorable site for peroxidative activity as described above (**Figure S5A**).
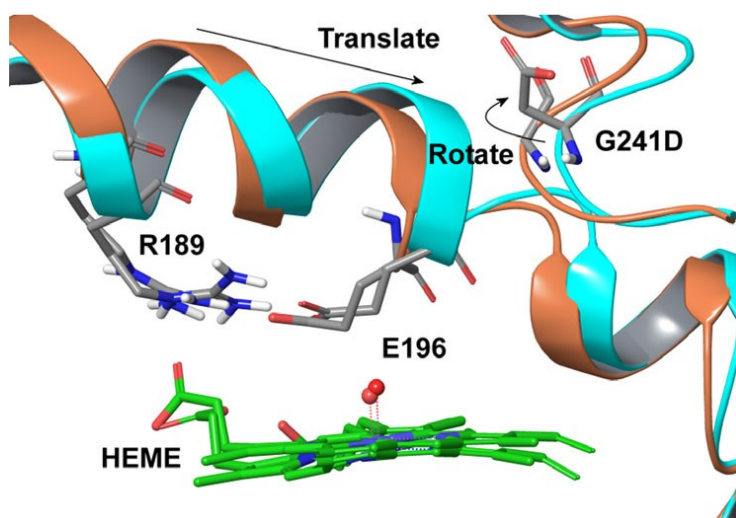

Figure 5. G241D-induced conformational changes. Heme binding site PELE simulations show that G241D leads to backbone rotation, exposing Asp241 to the solvent and avoiding the hydrophobic protein environment. This rotation shifts the loop containing Asp241 and creates space for the displacement of the α-helix hosting the catalytic Arg189–Glu196 pair. JaWa is shown in cyan and PaDa-I in orange, after α-carbon superposition of the entire protein.

Peroxygenase activity: To explain the effect of the two introduced mutations on k_{cat} and K_m, ligand exploration at the heme site was performed with PELE and use of the two AaeUPO1 variants JaWa and PaDa-I. Additionally, QM/MM calculations were carried out to investigate the mutations' effect on naphthalene oxygenation.

1) *Structural variations*: Heme binding site PELE simulations with naphthalene show that in the JaWa variant the G241D mutation induces a $\approx 20°$ rotation in the aspartate backbone dihedral (**Figure S5 B**), leading to a shift in the loop where it is located. Along with this rotation we observed the disruption of the hydrogen bond between Thr198 and Gly241 (highly stable in PaDa-I: 86 %

residence time versus 0% in JaWa). These changes create space for the concomitant displacement of the α-helix that hosts the catalytic acid-base (Arg189–Glu196) pair (**Figure 5**).

2) *Naphthalene binding*: The previously mentioned structural changes (induced by G241D), along with other minor variations in the protein's structure, caused a redistribution of the binding site occupation in JaWa mutant. Ideally, in order to facilitate epoxide intermediate formation, distances between C1 and C2 naphthalene carbon atoms and the heme catalytic oxygen should be below 3 Å for optimum reaction. As can be seen in Figure 6 A, analysis of naphthalene's binding site position during PELE simulations shows an increased population of optimum catalytic orientations in JaWa (shorter naphthalene–heme distances). This is in agreement with the 4.5-times decreased K_m value for this variant (**Table 2**).

3) *Hydroxylation of naphthalene*: Mutation-induced structural variations have repercussions not only for naphthalene's binding site orientation, but also for its hydroxylation. The α-helix structural changes induce a new alternative position for Arg189, closer to the substrate (**Figure 6B**). To compute the effect of this electrostatic environment change on naphthalene's oxidation we again turned to QM/MM calculations. Naphthalene ionization energies were computed for both PaDa-I and JaWa variants. The results show that the increase in the substrate's positive electrostatic environment increases the naphthalene's ionization energy (from 213.55 kcal mol^{-1} in PaDa-I to 214.57 kcal mol^{-1} in JaWa), leading to a higher activation barrier, as shown in previous studies[31,32] and in agreement with the lower k_{cat} value for JaWa observed experimentally.

On the other hand, the R257K mutation was not expected to affect the hydroxylation of naphthalene, but further PELE simulations with DMP as substrate show additional access to the active heme site for this compound, not present in naphthalene simulations, that could affect both K_m and k_{cat} for DMP. Moreover, it is known that some peroxidases have different inlets at the surface of the protein that serve for the one-electron oxidation of reducing substrates through a long-range electron-transfer pathway towards the heme domain, as de- scribed here for W24F variants.[33] The R257K substitution might affect some of these enzyme "circuits", possibly offering benefits in terms of improved thermostability through local rearrangements in secondary structures. Indeed, both the G241D and the R257K mutations modified the local B factor profiles (**Figure S6**).
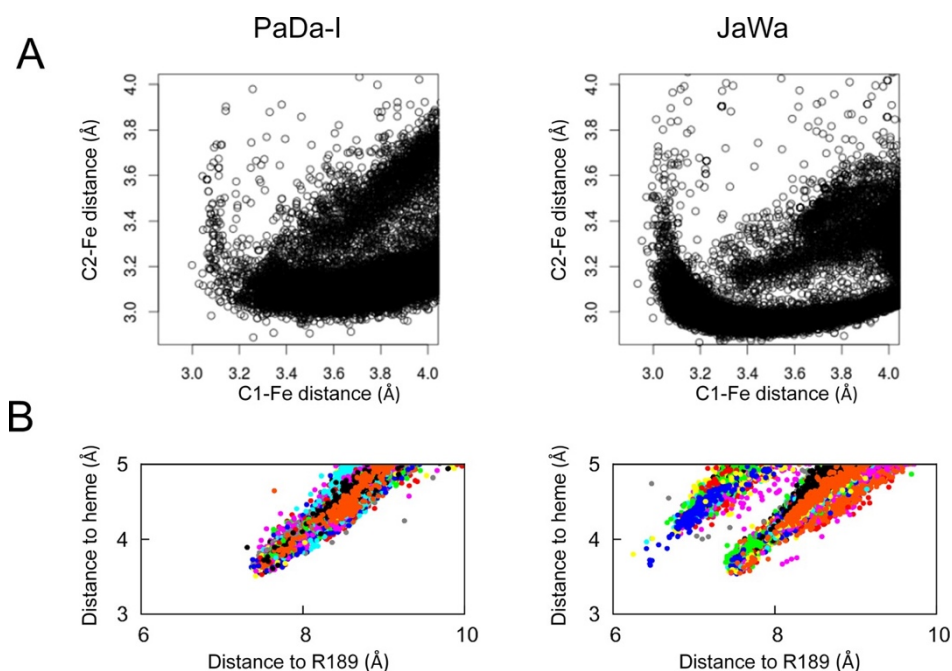
Figure 6. A) Distances between C1/C2 naphthalene carbons (hydroxylated carbons) and the heme catalytic oxygen showing a larger substrate population for optimum reactive distances ($<3\,\text{Å}$) in JaWa. B) Variation of naphthalene's center of mass (COM) to heme oxygen (compound I) versus naphthalene's COM to CZ atom in R189 in PaDa-I and JaWa during PELE simulations. The results show that in the case of JaWa the arginine side chain is on average closer to the substrate.

Conclusions

To date, over 1000 UPO-like genes from fungi have been mined from genomic databases; this highlights their wide- spread distribution in the fungal kingdom. Nevertheless, only three wild-type UPOs have been biochemically characterized and published (from *Coprinellus radians*, *Marasmius rotula*, and the above-mentioned *A. aegerita*). Some other UPOs have been identified (e.g., from *Chaetomium* sp.) but they have not yet been characterized in detail.[34] More recently, UPO genes from the genomes of *Coprinopsis cinerea* and from a soil mold were expressed heterologously.[21] The *Aae*UPO1 engineered here shows the highest selectivity and TTN for the production of 1- naphthol so far reported in this enzyme superfamily. Readily secreted in an active, soluble, and very stable form, this heter- ologous enzyme performs selective aromatic oxygenations in the absence of NAD(P)H cofactors and reductase domains. Its self-sufficient mono(per)oxygenase activity, along

with a dimin- ished peroxidative activity, make this UPO mutant a valuable biocat-alyst for future synthetic applications**.**

Experimental Section

Reagents and enzymes: ABTS, DMP, benzyl alcohol, 1-naphthol (naphthalen-1-ol), 2-naphthol (naphthalen-2-ol), 1,4-naphthoquinone, Fast Red TR Salt hemi(zinc chloride) salt, Taq DNA poly- merase, and the Yeast Transformation kit were purchased from Sigma–Aldrich. NBD was acquired from TCI America (Portland, OR, USA), and naphthalene from Acros Organics. The Genemorph II Random Mutagenesis kit (Mutazyme II) was obtained from Agilent Technologies, and the high-fidelity DNA polymerase *i*Proof was acquired from Bio-Rad. The BamHI and XhoI restriction enzymes were purchased from New England Biolabs and the protease-deficient *S. cerevisiae* strain BJ5465 from LGCPromochem (Barcelona, Spain). The Zymoprep Yeast Plasmid Miniprep kit and Zymoclean Gel DNA Recovery kit were from Zymo Research (Orange, CA, USA). The Nucleo-Spin Plasmid kit was purchased from MachereyNagel, and the oligonucleotides were synthesized by Isogen Life Science (Barcelona, Spain). All chemicals were re-agent-grade purity.

Laboratory evolution: The parent PaDa-I was obtained as described elsewhere.[23] After each round of directed evolution, the PCR products were loaded onto a pre-parative agarose gel and purified by use of the Zymoclean Gel DNA Recovery kit. The recovered DNA fragments were cloned under the control of the GAL1 pro-moter of the pJRoC30 expression shuttle vector, with use of BamHI and XhoI to linearize the plasmid and to remove the parent gene. The linearized vector was loaded onto a low-melting-point preparative agarose gel and purified with the Zy-moclean Gel DNA Recovery kit.

First generation : With use of PaDa-I as parent, error-prone PCR was carried out in a reaction mixture (50 μL final volume) containing di- methyl sulfoxide (DMSO, 3 %), RMLN (5'-cctctatactttaacgtcaagg-3', 0.37 μm), RMLC (5'-gcttacattcac-gccctccc-3', 0.37 μm), deoxynucleotide triphosphates (dNTPs, 0.8 mm, 0.2 mm each), Mutazyme II (Genemorph II kit, 0.05 U μL^{-1}), and the template (plasmid pJRoC30 containing PaDa-I, 2.8 μg, 300 ng of the target DNA). The mutagenic PCR was performed with a gradient thermocycler (Mycycler, BioRad, USA) and use of the following cycles: 95ºC for 2 min (1 cycle), 94ºC for 45 s, 53ºC for 45 s, and 74ºC for 3 min (28 cycles), and 74ºC for 10 min (1 cycle). The PCR products (200 ng) were mixed with the linearized plasmid (100 ng) and transformed into

competent *S. cerevisiae* cells for in vivo shuffling and cloning by using the Yeast Transformation kit. Transformed cells were plated on synthetic complete (SC) drop-out plates and incubated for 3 days at 30ºC. Grown colonies were selected and subjected to the dual high-throughput screening (HTS) assay and additional rescreenings, as described below.

Second generation : The best mutants obtained from the first generation were used to perform mutagenic StEP[35] combined with in vivo shuffling. The PCR reaction mixture contained DMSO (3 %), RMLN (90 n_M), RMLC (90 n_M), dNTPs (0.3 m_M, 0.075 m_M each), Taq DNA polymerase (0.05 U μL^{-1}), and the template (16 ng, pJRoC30 containing the four best mutants of the first generation). The PCR was performed with a gradient thermocycler and use of the following cycles: 95ºC for 5 min (1 cycle) and 94ºC for 30 s, 55 ºC for 20 s (90 cycles). The PCR products were mixed with the linearized vector (200 ng of each PCR product and 100 ng of linearized plasmid), transformed into competent *S. cerevisiae* cells and treated as for the first generation.

W24F site-directed variants: PCR reactions for each variant (PaDa-I and JaWa) were carried out using primers (F24FOR: 5'-ctcacccatttaagccgcttcgacctggcga-tattcgtggac-3' and F24REV: 5'-gtccacgaatatcgccaggtcgaagcggcttaaatgggtgag-3'; (mutations sites are underlined). The PCR conditions were: 1) 50 µL final volume, 3% DMSO, 0.5 µM RMLN, 0.5 µM F24REV, 1 mM dNTPs (0.25 mM each), 0.02 U/µL high fidelity DNA polymerase iProof and 10 ng of the templates; and (2) 50 µL final volume, 3% DMSO, 0.5 µM F24FOR, 0.5 µM RMLC, 1 mM dNTPs (0.25 mM each), 0.02 U/µL high fidelity DNA polymerase iProof and 10 ng of the tem-plates. The themocycler parameters were: (i) 98°C for 30 s (1 cycle), 98°C for 10 s, 47°C for 25 s, 72°C for 15 s (28 cycles), and 72°C for 10 min (1 cycle); or (ii) 98°C for 30 s (1 cycle), 98°C for 10 s, 58°C for 25 s, 72°C for 45 s (35 cycles), and 72°C for 10 min (1 cycle). PCR products belonging to each template (200 ng each) were mixed with the linearized vector (100 ng) and transformed into *S.cerevisiae* for *in vivo* gene reassembly and cloning by IVOE.[36]

High-throughput screening assay. Individual clones were picked and inoculated in sterile 96-well plates (Greiner Bio-One, GmbH, Germany), referred to as master plates, containing 200 µL of minimal expression medium per well [filtered yeast nitrogen base (100 mL, 6.7 %), filtered yeast synthetic drop-out medium supple-ment without uracil (100 mL, 19.2 g L^{-1}), filtered potassium phosphate buffer (pH 6.0, 67 mL, 1 m), filtered galactose (111 mL, 20 %), filtered MgSO4 (22 mL, 0.1 M), absolute ethanol (31.6 mL), chloramphenicol (1 mL, 25 gL^{-1}), and *dd*H_2O (to 1000 mL)]. In each plate, column 6 was inoculated with the corresponding parent,

and one well (H1-control) was inoculated with untransformed S. cerevisiae cells. The plates were sealed to prevent evaporation and incubated at 30 8C, 220 rpm, and 80 % relative humidity in a shaker (Minitron, INFORS, Switzerland) for five days. The master plates were centrifuged (Eppendorf 5810R centrifuge, Germany) for 10 min at 3,500 RPM and 4°C. 20 µL of supernatant were transferred from the master plates to two replica plates by using a liquid handler robotic station Freedom EVO (Tecan, Switzerland). 180 µL of reaction mixture with DMP or naphthalene were added to each replica plate with the help of a pippeting robot (Multidrop Combi Reagent Dispenser, Thermo Scientific, USA). DMP reaction mixture contained 100 mM potassium phosphate buffer at pH 7.0, 3 mM DMP and 1 mM H_2O_2 (at the same time, the same screening was carried out but with the addition of 10% of acetonitrile to assess possible changes in activity due to resistance to this organic co-solvent, needed for naphthalene solution). Reaction mixture with naphthalene contained 100 mM potassium phosphate buffer pH 7.0, 0.5 mM naphthalene, 10% acetonitrile and 1 mM H_2O_2. Plates were stirred briefly and the initial absorptions at 469 nm and 510 nm, respectively, were recorded in the plate reader (SPECTRA-Max Plus 384, Molecular Devices, Sunnyvale, CA). After a reaction time of 10 min, 20 µL of Fast Red (Fast Red TR Salt hemi(zinc chloride) salt) were added per well (final concentration of 0.5 mM). The plates were incubated at room temperature until a red (naphthalene-Fast Red) or orange (DMP) color developed, and the absorption was measured again. The values were normalized against the parental type in the corresponding plate. To rule out false positives, two re-screenings were carried out as reported elsewhere. Finally, a third re-screening was performed in order to assess kinetic stability (the re-screenings plus the kinetic stability (T_{50}) experiments are described in [23]).

Biochemical characterization. PaDa-I and JaWa variants were produced and purified as described in [23].

Steady-state kinetic constants. ABTS kinetic constants for UPO were estimated in sodium phosphate/citrate buffer (pH 4.0, 100 mM), containing H2O2 (2 mM), and for the rest of the substrates in potassium phosphate buffer (pH 7.0, 100 mm) containing H2O2 (2 mM, DMP) or H2O2 [1 mM, naphthalene, in acetonitrile (20 %)— final concentration]. For H_2O_2, benzyl alcohol was used as a reducing substrate under the corresponding saturated conditions. Reactions were performed in triplicate, and substrate oxidations were followed through spectrophotometric changes (ABTS, $\varepsilon_{418} = 36{,}000$ M^{-1} cm^{-1}; DMP, $\varepsilon_{469} = 27{,}500$ M^{-1} cm^{-1}; NBD, $\varepsilon_{425} = 9{,}700$ M^{-1} cm^{-1}, and naphthalene $\varepsilon_{303} = 2{,}010$ M^{-1} cm^{-1}, benzyl alcohol, $\varepsilon_{280} = 1{,}400$ M^{-1} cm^{-1}). Naphthalene kinetics were performed according to the protocol described elsewhere [37]. To calculate the K_m and k_{cat} values, the average V_{max} was represented

against substrate concentration and fitted to a single rectangular hyperbola function using SigmaPlot 10.0, where parameter *a* was equaled to k_{cat} and parameter *b* was equaled to K_m.

HPLC analysis. The reaction mixtures were analyzed by reversed-phase chromatography (HPLC) with equipment consisting of a tertiary pump (Varian/Agilent Technologies) coupled to an autosampler (Merck Millipore) and an ACE C18 PFP (pentafluorophenyl, 15 cm x 4.6 mm) column at 45ºC. Detection was performed with a PDA (Varian/Agilent Technologies). The mobile phase was methanol (70%) and *dd*H_2O (30 %, both with 0.1% acetic acid) at a flow rate of 0.8 mL.min^{-1}. The reaction was quantified at 268 nm (from HPLC standards). For the 15 min reaction, the mixture contained purified enzyme (6.6 nM), naphthalene (1 mM), acetonitrile (20 %), and H_2O_2 (1 mM) in potassium phosphate (pH 7.0, 100 mM, final volume of 1 mL). The reaction was started by the addition of the H_2O_2 and stopped by addition of HCl (20 µL, 37%); a sample (10 µL) was injected and analyzed. For the longer reactions, the conditions were as described above, but the reaction was not stopped with HCl. A sample (10 µL) was injected and analyzed at different times (1 to 270 min). To determine the kinetic values for 1-naphthol, the reaction was performed with the pure enzymes (40 nM), 1-naphthol (1 mM), acetonitrile (20 %), and H2O2 (1 mM) in potassium phosphate (pH 7.0, 100 mM, final volume of 0.2 µL). Standard deviations were lower than 5% in all experiments.

MALDI–TOF-MS analysis and pI determination. Experiments were performed on an Autoflex III MALDITOF/TOF instrument with a smartbeam laser (Bruker Daltonics). A laser power just above the ionization threshold was used in order to acquire the spectra and the samples were evaluated in the positive-ion detection. External calibration was performed, using the BSA from Bruker, covering the range 15,000 – 70,000 Da. To determine UPO's pI, 8 µg of purified enzyme were subjected to two-dimensional electrophoresis gel. These determinations were carried out at the Proteomic and Genomic Services from CIB (CSIC, Spain).

LC/MS. A mass spectrometer with hybrid mass analyzer Q-TOF (QSTAR, AB-Sciex, MA, USA) was used. The ionization source used was electrospray (ESI), with methanol as ionization phase. The inlet system used was direct injection in a HPLC 1100 (Agilent Technologies, USA). The resolution of the assay was 9,000 FWHM (full width at half maximum), the accurateness was 5-10 ppm and it was carried out in negative mode.

DNA sequencing. Plasmid-containing variant *upo*1 genes were sequenced by using an ABI 3730 DNA Analyzer/Applied Biosystems Automatic Sequencer from

Secugen (Spain). The primers used were: RMLN; apo1secdir (gaaggcgacgccag-tatgacc); apo1secrev (ggtcatactggcgtcgccttc) and RMLC.

Computational analysis

System preparation for molecular modeling: The structure of wild- type UPO1 (purified from *A. aegerita* culture) at a resolution of 2.1 Å (PDB ID: 2YOR) was used as departure point for modeling of the PaDa-I and JaWa variants.[27] The wild-type UPO1 crystal was used. Five different mutations (V57A-L67F-V75I-I248V-F311L) were introduced into the model PaDa-I variant, with two additional mutations (G241D-R257K) for JaWa. Because the optimal pH for DMP and naphthalene activity is ≈ 7, mutated structures (and the UPO1 crystal) were prepared accordingly with the aid of the Schrödinger Protein Preparation Wizard and the H+ + web server.[38, 39] All acidic residues were deprotonated except Asp85. Histidines were δ-protonated, with the exceptions of His82 (ε-protonated) and His118 and His251 (double-protonated). To relax the systems after mutation, and to investigate the mutations' possible effects on the protein structure, UPO1 variants were subjected to 5 ns molecular dynamics (MD) with Desmond.[40] Finally, the heme site was modeled as compound I after being fully optimized in the protein environment with quantum mechanics/molecular mechanics (QM/MM) and use of QSite.[41] DMP and naphthalene molecules were also optimized with Jaguar[42] at the DFT/M06 level with the 6–31G** basis and PBF implicit solvent in order to obtain their electrostatic potential atomic charges.

Protein energy landscape exploration (PELE) computational analysis: Once PaDa-I and JaWa structures had been prepared and ligands optimized, heme binding site and global protein surface exploration were performed with PELE, a Monte Carlo algorithm capable of effectively sampling the protein–ligand conformational space.[38] For the protein surface exploration, ligands were placed manually in 20 initial random positions on the protein's surface, and the ligand was allowed to explore freely. For each system, 160 independent 48 h simulations were performed. On the other hand, for the heme binding site exploration, the substrates were placed manually in identical positions at the entrance of the heme-access channel. From there, ligands were spawned inside the protein by PELE. Once the ligand's center of mass was within 5 Å of the heme catalytic oxygen, it was free to explore the active site cavity with 96 independent 48 h simulations.

QM/MM simulations: Hybrid quantum mechanics/molecular mechanics (QM/MM) calculations were carried out for two different purposes: to identify the

most oxidizable surface residue in UPO1 protein, and to investigate the mutations' effect on naphthalene oxidation. For the first purpose, a QM/MM calculation was per- formed in wild-type UPO1 by including all potential surface oxidation sites (nine tyrosine residues and one tryptophan) in the quantum region, subtracting one electron, and computing the spin density. A subsequent QM/MM pairwise compar- ison was performed with residues that showed a clear preference for oxidation (Trp24 and Tyr47). Calculations were performed at the DFT M06L(lacvp*)/ OPLS level. On the other hand, to study the differences between PaDa-I and JaWa naph- thalene oxygenation, five ionization energies of the substrate (located on the bind- ing site of the variants) were computed. Energies were taken from single-point cal- culations by using the B97-D3(cc-pVTZ(-f) + +) basis set. All QM/MM calcula- tions were performed with Qsite.

Acknowledgements

Keywords: 1-naphthol · directed evolution · enzyme catalysis ·peroxi- dative activity · peroxygenases

[1] G. Booth in *Ullmann's Encyclopedia of Industrial Chemistry, Vol. 23*, Wiley-VCH, Weinheim, 2000, pp. 671 – 723.

[2] K. R. Jegannathan, P. H. Nielsen, *J. Cleaner Prod.* 2013, *42*, 228 – 240.

[3] K. Kudo, T. Ohmae, A. Uno (Sumitomo Chemical Company), United States Patent 3,935,282, 1976.

[4] L. Schuster, B. Seid (BASF AG), United States Patent, 4,171,459, 1979.

[5] I. Calinescu, R. Avram, *Rev. Chem.* 1994, *45*, 97 – 103.

[6] I. Calinescu, R. Avram, H. Iovu, *Rev. Chem.* 1994, *45*, 299 – 305.

[7] I. Calinescu, R. Avram, *Rev. Chem.* 1994, *45*, 865 – 867.

[8] J. B. van Beilen, W. A. Duetz, A. Schmid, B. Witholt, *Trends Biotechnol.*

2003, *21*, 170 – 177.

[9] B. Buhler, A. Schmid, *J. Biotechnol.* 2004, *113*, 183 – 210.

[10] V. B. Urlacher, R. D. Schmid, *Curr. Opin. Chem. Biol.* 2006, *10*, 156 – 161.

[11] R. Ullrich, M. Hofrichter, *Cell. Mol. Life Sci.* 2007, *64*, 271 – 293.

[12] H. Joo, Z. Lin, F. H. Arnold, *Nature* 1999, *399*, 670 – 673.

[13] P. C. Cirino, F. H. Arnold, *Angew. Chem. Int. Ed.* 2003, *42*, 3299 – 3301; *Angew. Chem.* 2003, *115*, 3421 – 3423.

[14] P. Meinhold, M. W. Peters, A. Hartwick, A. R. Hernandez, F. H. Arnold, *Adv. Synth. Catal.* 2006, *348*, 763 – 772.

[15] P. S. Coelho, E. M. Brustad, A. Kannan, F. H. Arnold, *Science* 2013, *339*, 307 – 310.

[16] K. A. Canada, S. Iwashita, H. Shim, T. K. Wood, *J. Bacteriol.* 2002, *184*, 344 – 349.

[17] L. Rui, Y. M. Kwon, A. Fishman, K. F. Reardon, T. K. Wood, *Appl. Environ. Microbiol.* 2004, *70*, 3246 – 3252.

[18] J. Garikipati, A. M. McIver, T. L. Peeples, *Appl. Environ. Microbiol.* 2009, *75*, 6545 – 6552.

[19] R. Ullrich, J. Nuske, K. Scheibner, J. Spantzel, M. Hofrichter, *Appl. Environ. Microbiol.* 2004, *70*, 4575 – 4581.

[20] M. Hofrichter, R. Ullrich, M. J. Pecyna, C. Liers, T. Lundell, *Appl. Microbiol. Biotechnol.* 2010, *87*, 871 – 897.

[21] M. Hofrichter, R. Ullrich, *Curr. Opin. Chem. Biol.* 2014, *19*, 116 – 125.

[22] M. Kluge, R. Ullrich, C. Dolge, K. Scheibner, M. Hofrichter, *Appl. Microbiol. Biotechnol.* 2009, *81*, 1071 – 1076.

[23] P. Molina-Espeja, E. Garcia-Ruiz, D. Gonzalez-Perez, R. Ullrich, M. Hof- rich- ter, M. Alcalde, *Appl. Environ. Microbiol.* 2014, *80*, 3496 – 3507.

[24] P. Molina-Espeja, S. Ma, D. M. Mate, R. Ludwig, M. Alcalde, *Enzyme Microb. Technol.* 2015, *73 – 74*, 29 – 33.

[25] H. Zollinger, *Diazo Chemistry I*, Wiley-VCH, Weinheim 2004, pp. 1 – 13.

[26] E. Aranda, R. Ullrich, M. Hofrichter, *Biodegradation* 2010, *21*, 267 – 281.

[27] K. Piontek, E. Strittmatter, R. Ullrich, G. Grçbe, M. J. Pecyna, M. Kluge, K.

Scheibner, M. Hofrichter, D. A. Plattner, *J. Biol. Chem.* 2013, *288*, 34767 – 34776.

[28] A. Madadkar-Sobhani, V. Guallar, *Nucleic Acids Res.* 2013, *41*, W322 – W328.

[29] V. Guallar, F. H. Wallrapp, *Biophys. Chem.* 2010, *149*, 1 – 11.

[30] D. Linde, R. Pogni, M. CaÇellas, F. Lucas, V. Guallar, M. C. Baratto, A. Sini- cropi, V. Saez-Jimenez, C. Coscolfn, A. Romero, F. J. Medrano, F. J. Ruiz- DueÇas, A. T. Martfnez, *Biochem. J.* 2015, *466*, 253 – 262.

[31] D. Kumar, B. Karamzadeh, G. N. Sastry, S. P. de Visser, *J. Am. Chem. Soc.*

2010, *132*, 7656 – 7667.

[32] A. J. M. Ribeiro, D. Santos-Martins, N. Russo, M. J. Ramos, P. A. Fer- nandes, *ACS Catal.* 2015, *5*, 5617 – 5626.

[33] B. Valderrama, M. Ayala, R. Vazquez-Duhalt, *Chem. Biol.* 2002, *9*, 555 – 565.

2015, *851*, 341 – 368.

[34] H. Zhao, L. Giver, Z. Shao, J. A. Affholter, F. H. Arnold, *Nat. Biotechnol.*

1998, *16*, 258 – 261.

[35] M. Alcalde, *Methods Mol. Biol.* 2010, *634*, 3 – 14.

[36] M. G. Kluge, R. Ullrich, K. Scheibner, M. Hofrichter, *Appl. Microbiol. Bio- technol.* 2007, *75*, 1473 – 1478.

[37] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, *J. Comput.-Aided Mol. Des.* 2013, *27*, 221 – 234.

[38] R. Anandakrishnan, B. Aguilar, A. V. Onufriev, *Nucleic Acids Res.* 2012, *40*, W537 – W541.

[39] *Desmond Molecular Dynamics System*, version 2.2, D. E. Shaw Research, New York, NY, 2009. Maestro-Desmond Interoperability Tools, ver- sion 2.2, Schrçdinger, New York.

[40] QSite, version 5.7, Schrçdinger, LLC, New York, 2011.
Jaguar, version 8.1, Schrçdinger, LLC, New York, 2013.

CHEMBIOCHEM

Supporting Information

Synthesis of 1-Naphthol by a Natural Peroxygenase Engineered by Directed Evolution

Patricia Molina-Espeja,[a] Marina Cañellas,[b] Francisco J. Plou,[a] Martin Hofrichter,[c] Fatima Lucas,[b, d] Victor Guallar,[d, e] and Miguel Alcalde*[a]

cbic_201500493_sm_miscellaneous_information.pdf
cbic_201500493_sm_movie.mp4

Figure S1. Directed evolution of *Aae*UPO1. From rounds 1 to 5 the enzyme was improved for functional expression and activity (the accumulated mutations are shown as grey squares).[23] Departing from the PaDa-I variant, two further rounds of evolution were performed to improve 1-naphthol synthesis (new mutations are shown as black squares). The signal peptide is represented in dark red and the mature protein in red. Thermostability (T_{50}) was estimated from the culture supernatants and the activity (given in %) was measured from *S. cerevisiae* microcultures in 96-well plates (second re-screening); n.m. not measurable.

Figure S2. Biochemical characteristics of UPO variants. A) Spectroscopic features of PaDa-I (thin line) and JaWa (thick line) in the resting state; AU, arbitrary units. B) Thermostability (T_{50}) of PaDa-I (black circles) and JaWa (white circles). Experiments were performed on fresh supernatants, and each point represents the mean and standard deviation of 3 different experiments. C) Stability of PaDa-I (black bars) and JaWa (grey bars) at high concentrations of acetonitrile. Stability was assessed after a 5 h incubation at 20°C in 100 mM potassium phosphate buffer pH 7.0 containing increasing concentrations of acetonitrile (50 to 100%). At the end of the incubation, aliquots were removed and analyzed with the ABTS assay (2 mM H_2O_2 and 0.3 mM ABTS in 100 mM sodium phosphate/citrate buffer pH 4.4). The error bars indicate the standard deviations.

Figure S3. Mass spectrometry analysis of the reaction products. The reactions mixtures with A) PaDa-I or B) JaWa contained 20 nM of purified enzyme, 1 mM naphthalene, 20% acetonitrile and 1 mM H_2O_2 in 100 mM potassium phosphate pH 7.0. After 250 min, the reaction products were analyzed by electrospray ionization in negative mode (ESI-) with methanol as the ionization phase. The major peak represents naphthols (both 1-naphthol and 2-naphthol are equally charged upon ionization) with a $[M-H]^-$ value of 143.05 m/z. Minor peaks appeared to be identical for using PaDa-I or JaWa, which agrees with the m/z expected for the buffer used.

Figure S4. Protein modeling of A) PaDa-I and B) JaWa with PyMOL Molecular Graphics System (Version 1.3 Schrödinger, LLC) based on the *Aae*UPO1 crystal structure (PDB accession number 2YOR). For PaDa-I: the new mutations relative to the native UPO are shown in dark green and underlined, and the residues in yellow and underlined in zigzag are those that are changed in JaWa. For JaWa: the new mutations are depicted in orange and underlined in zigzag. The heme is shown in CPK coloring; Fe^{3+} in red; structural Mg^{2+} in salmon; the Cys36 axial heme ligand and the disulfide bridge formed between Cys278 and Cys319 in light green; the *cis*-peptide bond between P108 and P109 is in brown; the 5 Phe residues that mediate the adjustment of the substrates, Phe 69, Phe 76, Phe 121, Phe 191 and Phe 199, are in pink; and the two catalytic residues, R189 and E196, are in light cyan.

Figure S5. A) QM/MM spin density distribution on Trp24 and Tyr47 in *Aae*UPO1 when including both residues in the quantum region. B) Variation of the 241 residue dihedral during PELE simulations. CB-CA-C-O dihedral was measured for JaWa (cyan) and HA3-CA-C-O for PaDa-I (orange). The average value for JaWa is -60.2° and that for PaDa-I is -43.6°.

Figure S6. B-factors for evolved UPOs. B-factor representation of the parental (PaDa-I, left) and JaWa variant (right), obtained using the PyMOL Molecular Graphics System (Version 1.3 Schrödinger, LLC). A) Detail of the mutation at position 257, located at the surface. Flexibility increases in the following color code order: Blue<green<yellow<red. B) Putty representation of the full UPO structure (the thicker the region, the more flexible it is).
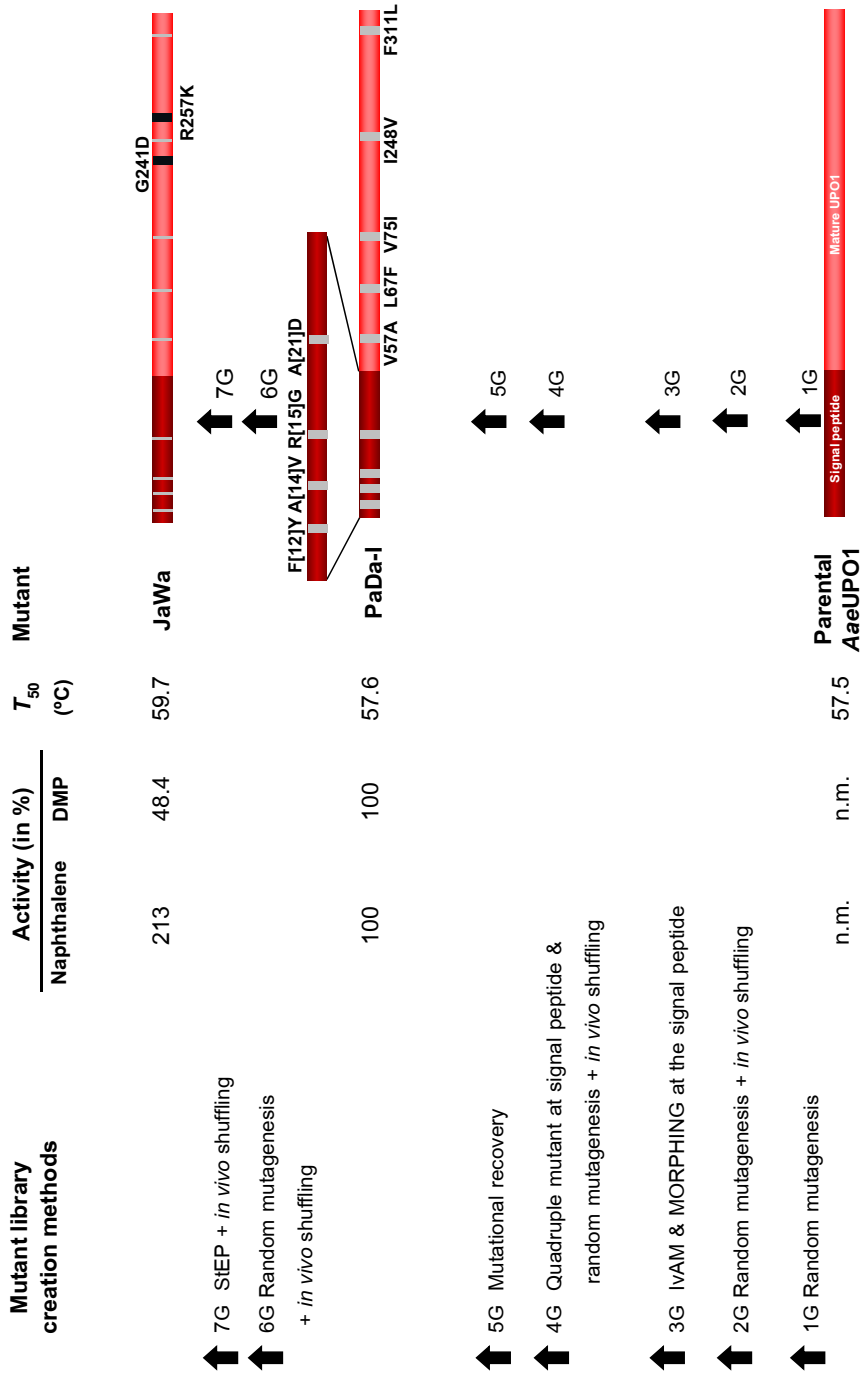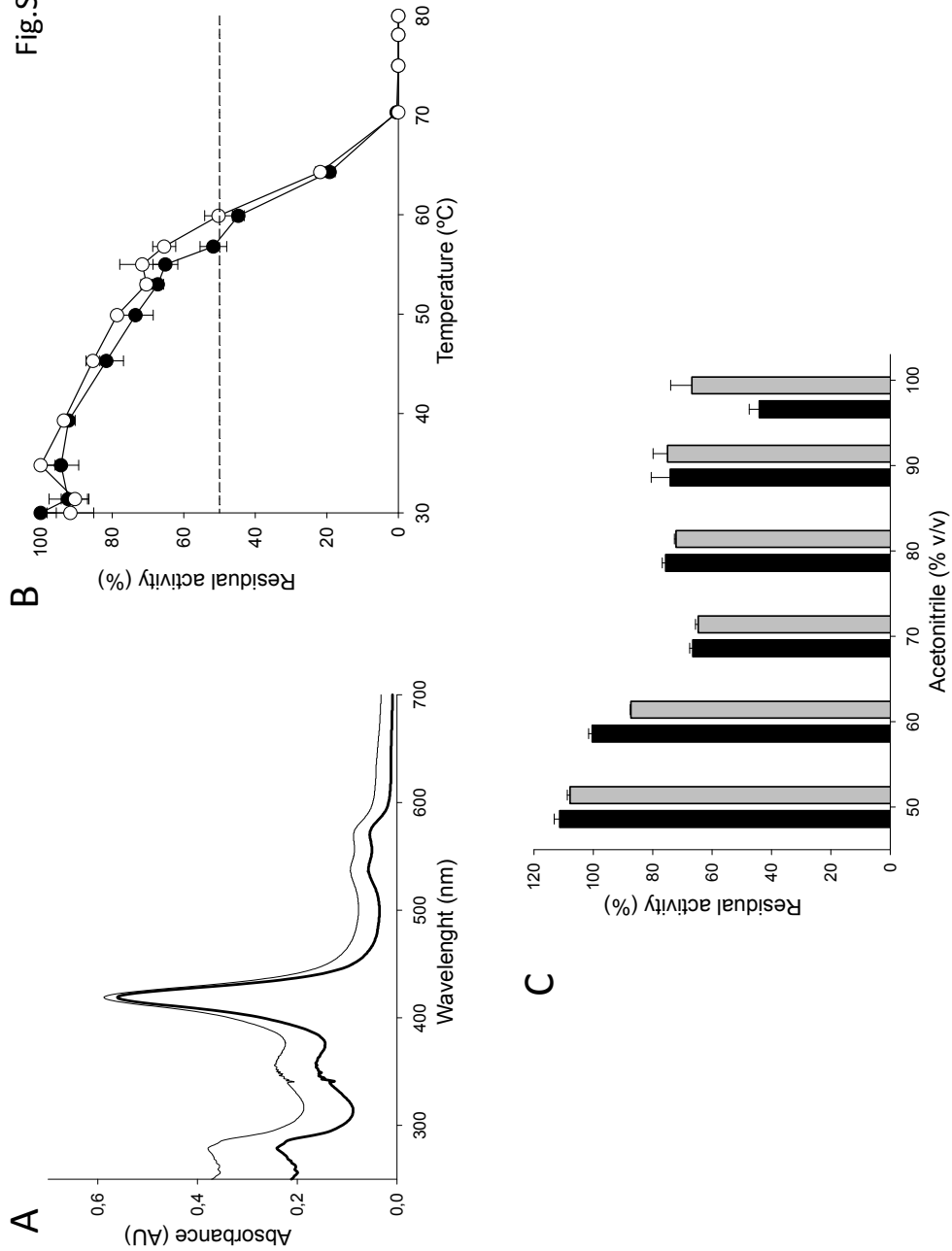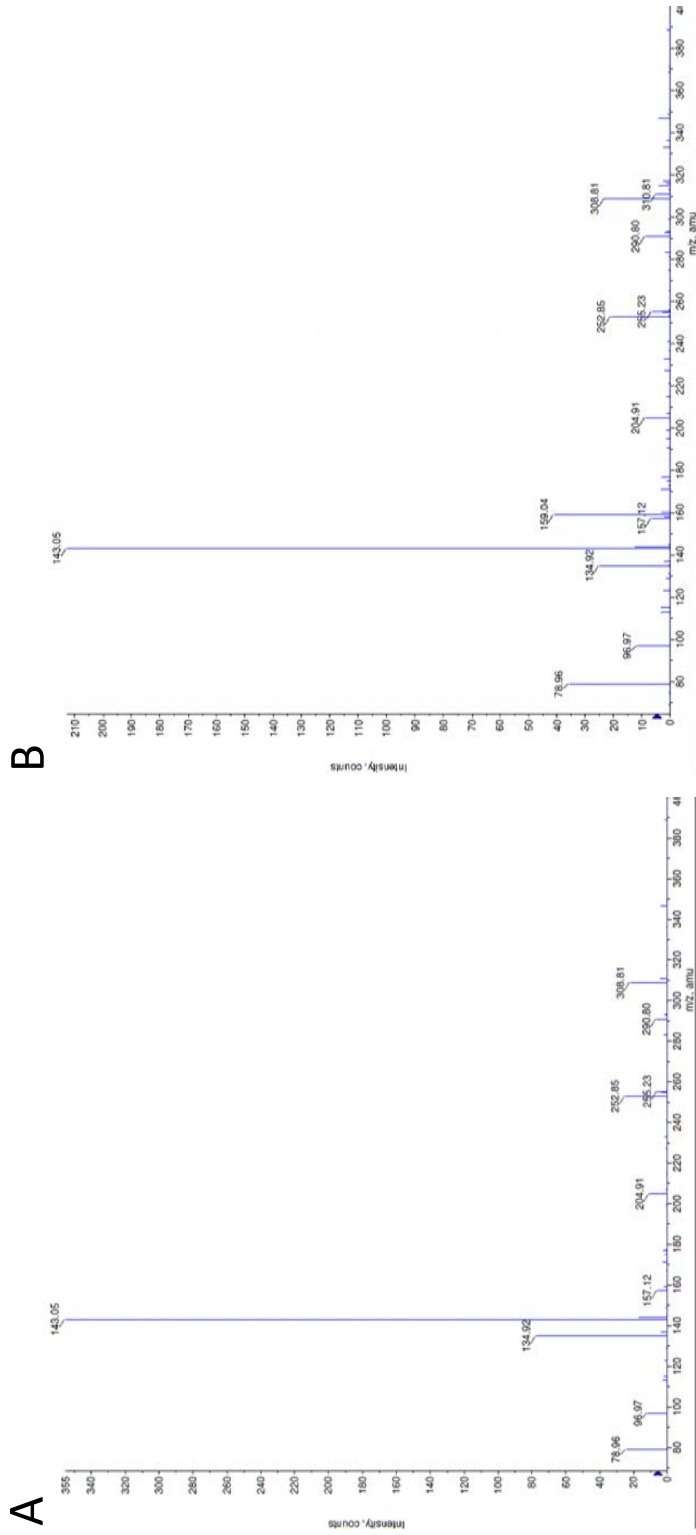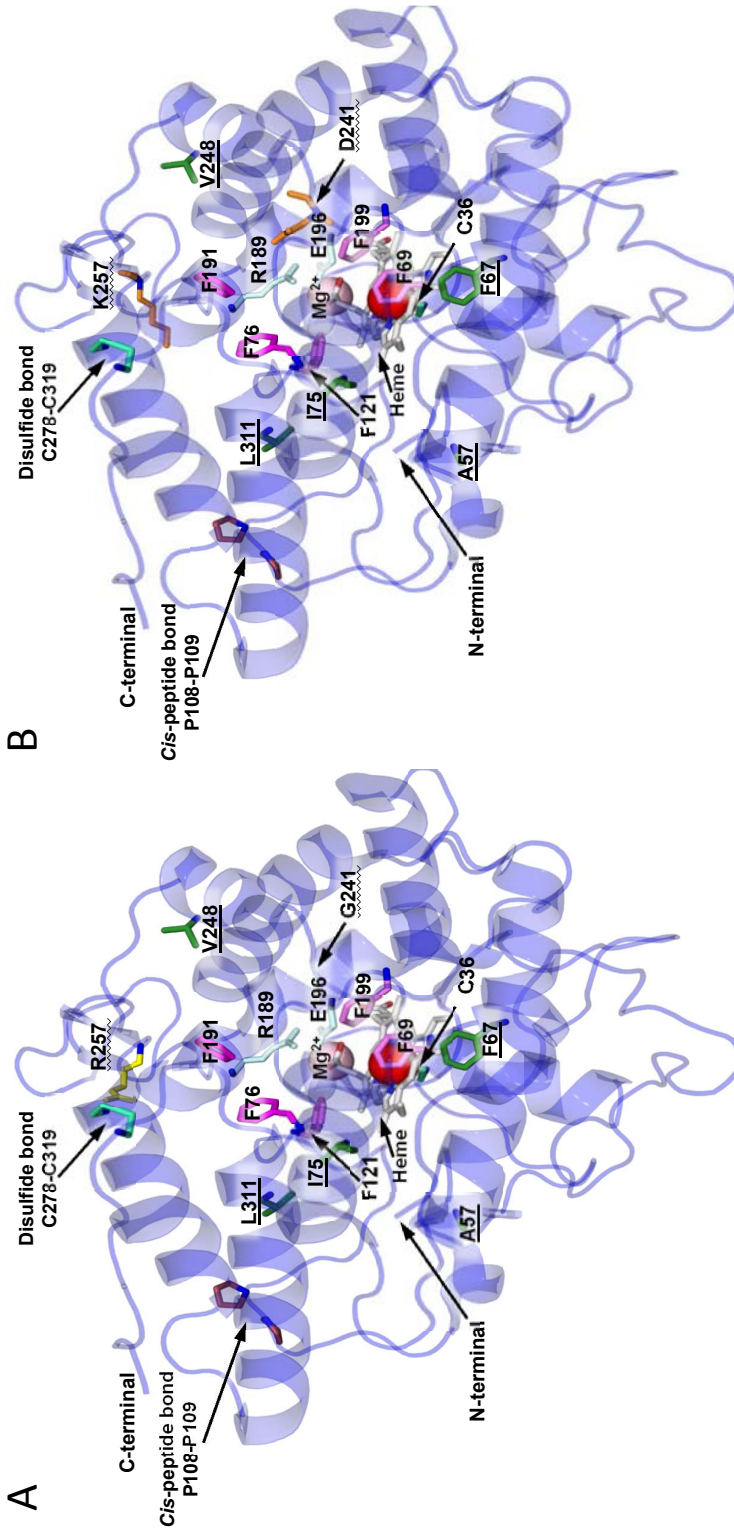
Fig.S1

Fig.S2

Fig. S3

Fig. S4

Fig. S5

B

A

Fig. S6

3.2.3 *Article* **VI: Selective synthesis of the human drug metabolite 5´-hydroxypropranolol by an evolved self-sufficient peroxygenase.**

Patricia Gomez de Santos[†], <u>Marina Cañellas</u>[‡], Florian Tieves[§], Sabry H. H. Younes[§], Patricia Molina-Espeja[†], Martin Hofrichter[φ], Frank Hollmann[§], Victor Guallar[‡] and Miguel Alcalde[†*]

[†] Department of Biocatalysis, Institute of Catalysis, CSIC, 28049 Madrid, Spain.

[‡] Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, 08034 Barcelona, Spain.

[§] Department of Biotechnology, Delft University of Technology, van der Massweg 9, 2629HZ Delft, The Netherlands.

[φ] Department of Bio- and Environmental Sciences, TU Dresden, International Institute Zittau, Mark 23, 02763 Zittau, Germany.
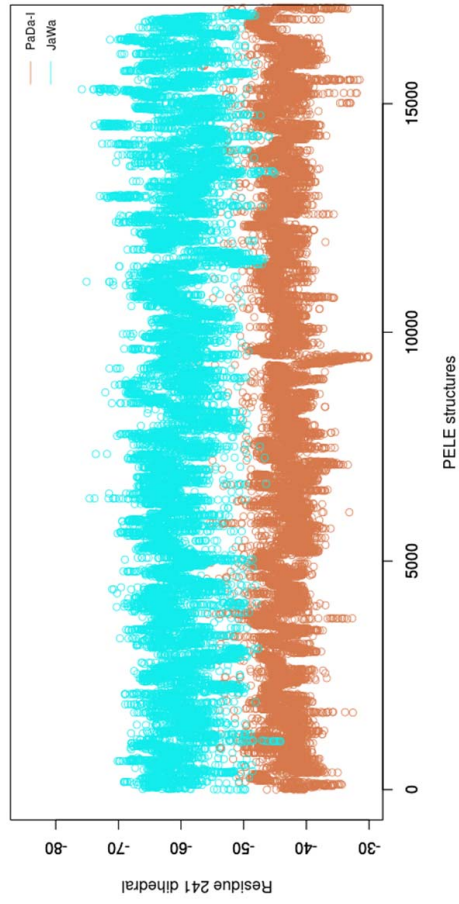
* Correspondence should be addressed to: M.A. (<u>malcalde@icp.csic.es</u>).
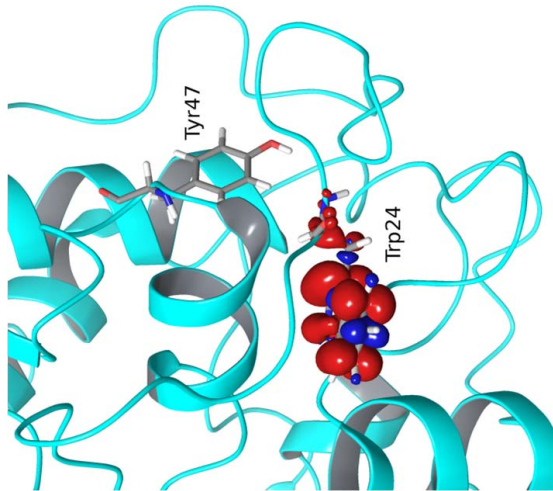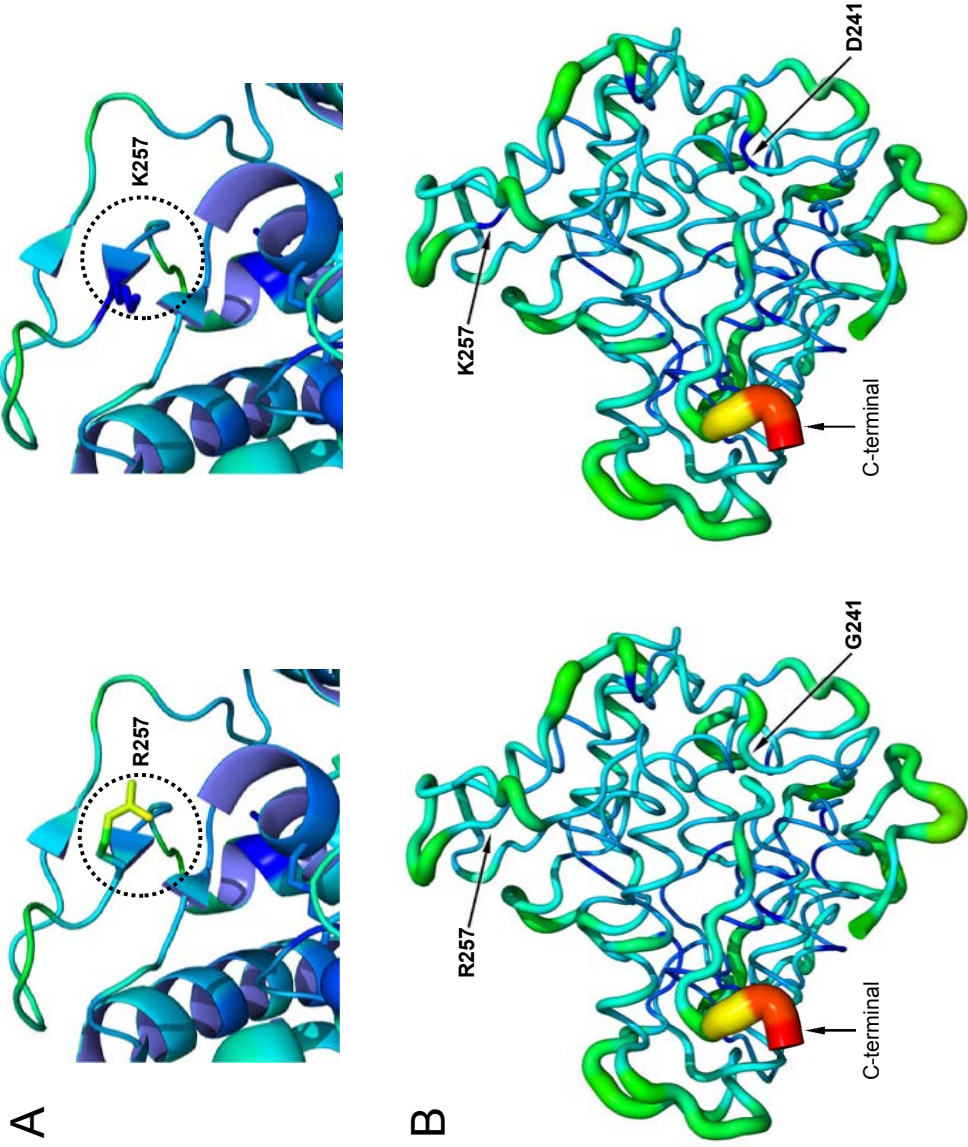
Selective synthesis of the human drug metabolite 5´-hydroxypropranolol by an evolved self-sufficient peroxygenase

Patricia Gomez de Santos[†], Marina Cañellas[‡], Florian Tieves[§], Sabry H. H. Younes[§], Patricia Molina-Espeja[†], Martin Hofrichter[φ], Frank Hollmann[§], Victor Guallar[‡] and Miguel Alcalde[†*]

[†]Department of Biocatalysis, Institute of Catalysis, CSIC, 28049 Madrid, Spain.

[‡]Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, 08034 Barcelona, Spain.

[§]Department of Biotechnology, Delft University of Technology, van der Massweg 9, 2629HZ Delft, The Netherlands.

[φ]Department of Bio- and Environmental Sciences, TU Dresden, International Institute Zittau, Mark 23, 02763 Zittau, Germany.

* Correspondence should be addressed to: M.A. (malcalde@icp.csic.es).

ABSTRACT

Propranolol is a widely used beta-blocker that is metabolized by human liver P450 monooxygenases into equipotent hydroxylated human drug metabolites (HDMs). It is paramount for the pharmaceutical industry to evaluate the toxicity and activity of these metabolites but unfortunately, their synthesis has hitherto involved the use of severe conditions, with poor reaction yields and unwanted by-products. Unspecific peroxygenases (UPOs) catalyze the selective oxyfunctionalization of C-H bonds and they are of particular interest in synthetic organic chemistry. Here, we describe the engineering of UPO from *Agrocybe aegerita* for the efficient synthesis of 5´-hydroxypropranolol (5´-OHP). We employed a structure-guided evolution approach combined with computational analysis, with the aim of avoiding unwanted phenoxyl radical coupling without having to dope the reaction with radical scavengers. The evolved biocatalyst showed a catalytic efficiency enhanced by two orders of magnitude and 99% regioselectivity for the synthesis of 5´-OHP. When the UPO mutant was combined with an H_2O_2 *in situ* generation system using methanol as sacrificial electron donor, total turnover numbers of up to 264,000 were achieved, offering a cost-effective and readily scalable method to rapidly prepare 5´-OHP.

KEYWORDS: unspecific peroxygenase, human drug metabolites, 5´-hydroxypropranolol, directed evolution, peroxygenative activity, peroxidative activity, *in situ* H_2O_2 generation system.

INTRODUCTION

New drugs are currently being designed thanks to a better understanding of the biological targets associated with different diseases. As such, modern organic chemistry is becoming more and more involved in the discovery and testing of new bioactive compounds.[1] The human liver is in charge of the metabolism of most drugs, principally through the catalytic action of cytochrome P450-monooxygenases (P450s). Their activities are responsible for the release of human drug metabolites (HDMs), which may be biologically active through different pharmacological, toxicological or physiological interactions. Thus, it is important to be capable of synthesizing HDMs in order to perform adequate drug bioavailability, pharmacodynamics and pharmacokinetics studies.[2,3] Indeed, the US Food and Drug Administration (FDA) guidelines for metabolites in safety testing (MIST) declare that all metabolites generated at >10% of the total parent drug-related exposure must be subjected to safety testing.[4] Among the most important HDMs are those derived from propranolol (1-naphthalen-1-yloxy-3-(propan-2-ylamino)propan-2-ol), a β-blocker widely used to treat high blood pressure, to control heart rhythm or to prevent migraines.[5] The two main hydroxylated metabolites of propranolol are 5´-hydroxypropranolol (5´-OHP) and 4´- hydroxypropranolol (4´-OHP) which have equipotent β-receptor antagonist activity compared to that of propranolol.[6] However, their chemical synthesis is associated with poor reaction yields, an excessive number of steps and it requires high-energy input and harsh conditions, limiting their utility.[7,8]

The use of enzymes to hydroxylate propranolol has been studied, such as human P450s isolated from hepatic microsomes or produced in heterologous hosts, yet the constraints on expression, instability and the low reaction rates are still serious obstacles that must be overcome.[9] Human P450s are membrane bound and cofactor dependent biocatalysts, such that a simpler and more autonomous system would be desirable. Accordingly, soluble bacterial P450-BM3 was engineered to work via the "peroxide shunt" pathway, *i.e.* fueled by catalytic amounts of H_2O_2 in the absence of redox cofactors (NADPH) and auxiliary flavoproteins, just like an "artificial" peroxygenase. In addition, an ensemble of P450-peroxygenase variants was tailored to transform propranolol into a complex mixture of compounds enriched in the dealkylation product desisopropylpropranolol (DIP), along with low amounts of 4´-OHP and 5´-OHP.[10] These experiments pointed the way to develop

a novel approach for the synthesis of HDMs using the first, "truly natural", unspecific peroxygenase (UPO; EC 1.11.2.1), an extracellular heme-thiolate enzyme with mono(per)oxygenase activity that was discovered at the beginning of this century.[11,12] Unlike P450s, the major product of UPO was shown to be 5´-OHP (91% regioselectivity), along with only minor amounts of DIP.[13] Despite this promising performance, UPOs can still not be used on an industrial scale due to the coexistence of peroxygenative (*P*; O-transferring two-electron oxidations) and peroxidative activities (*p*; one-electron oxidation reactions) in the same biocatalyst.[13-15] Hence, the enhancement of the *P:p* ratio is fundamental to perform C-H oxyfunctionalization with UPO, as demonstrated in recent works.[16,17] Indeed, the products of this peroxygenative activity on aromatics like propranolol, become substrates of UPO's peroxidative activity, which ultimately leads to the formation of a complex mixture of phenoxyl radicals (including semiquinones) and their disproportionation (quinones) and coupling products. This mixture is further clouded by non-enzymatic polymerization, affecting the final yields and the purification of the target compounds.[18,19] Such problems might be partially circumvented by doping the reaction with expensive phenoxyl radical scavengers (*e.g.* ascorbic acid), although this solution is far from practical in terms of developing a cost-effective strategy (**Scheme 1**).
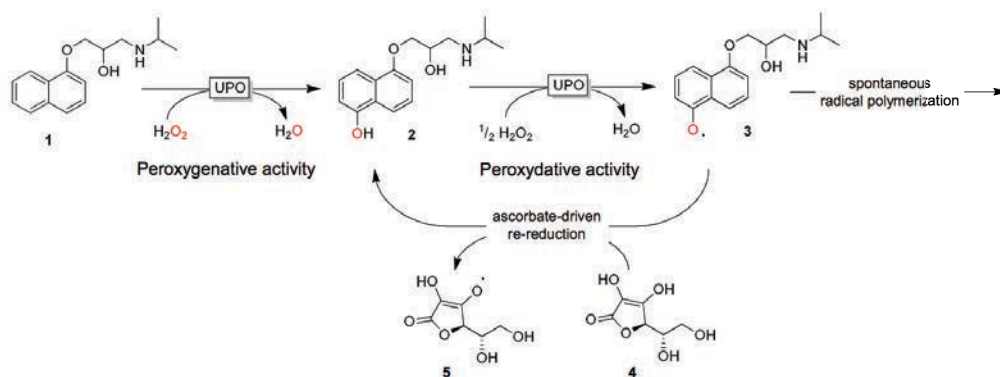
Scheme 1. Transformation of propranolol by UPO in the presence of ascorbic acid as a radical scavenger. Peroxygenative activity converts propranolol (**1**) into 5´-OHP (**2**), a substrate of the peroxidative activity of the enzyme that leads to the formation of phenoxyl radicals (**3**), which can in turn undergo non-enzymatic coupling and polymerization. The inclusion of ascorbic acid (**4**) in the reaction mixture can alleviate the formation of unwanted side-products by reducing the conversion of (**3**) to (**2**) while it is oxidized into ascorbyl radical (**5**).

In a previous study, we described the heterologous functional expression of UPO from the agaric basidiomycete *Agrocybe aegerita* (*Aae*UPO) in different yeasts via directed evolution.[20,21] This evolution platform (PaDa-I, secretion mutant carrying 4 mutations in the signal peptide and 5 more in the mature protein) enabled us to perform further rounds of random mutation and screening to tailor an efficient UPO for the synthesis of 1-naphthol, notably increasing the *P:p* ratio of the final JaWa variant (containing the G241D-R257K mutations).[16] In this work, we have combined computational simulations with directed evolution methods to design a UPO variant for the high-efficient, selective synthesis of 5´-OHP from propranolol in the absence of radical scavengers. We first made a benchmarking with native and evolved UPOs in the transformation of propranolol, selecting JaWa variant as the most suitable departure point for engineering. Focused mutagenesis and DNA recombination in yeast was coupled to a specific high-throughput assay for the synthesis of 5´-OHP, yielding a final variant that was characterized biochemically in depth. To bypass the oxidative inactivation suffered by UPO and maximize the total turnover numbers (TTN), an enzymatic cascade for the *in situ* supply of H_2O_2 using methanol as sacrificial electron donor was coupled to the reaction. The effect of the *P:p* ratio during the synthesis of 5´-OHP by the evolved UPO, was explained in depth by ligand migration and molecular dynamics simulations aimed at obtaining an atomic explanation of the mutations involved in the catalysis.

RESULTS & DISCUSSION

As a starting point for the present study, we used a colorimetric screening assay based on 4-aminoantypirine (4-AAP)[22] to benchmark the wildtype UPO (*Aae*-UPO), PaDa-I and JaWa mutants in the reaction with propranolol, assessing the variants in terms of their activity in the presence and absence of ascorbic acid (AA); **Figures S1, S2 and S3 in the Supporting Information**. Significantly, the JaWa mutant outperformed both the *Aae*UPO and PaDa-I variants by ~4-fold (both the presence and absence of AA; **Figure S3b in the Supporting Information**). Given that JaWa showed TTN of 50,000 in the transformation of naphthalene into 1-naphthol,[16] the structural similarities between naphthalene and propranolol, and the location of the G241D mutation at the entrance of the heme access channel, we searched for new catalytic motifs to be subjected to focused evolution. Molecular docking simulations[23] with propranolol underlined the possibility that along with Phe191, the G241D mutation favored the anchoring of propranolol (**Figure 1**). Accordingly, we selected the D187-V248 segment for random mutagenesis and DNA recombination by MORPHING (mutagenic organized recombination process by

homologous in vivo grouping),[24] excluding the introduction of destabilizing mutations in the remaining protein structure.
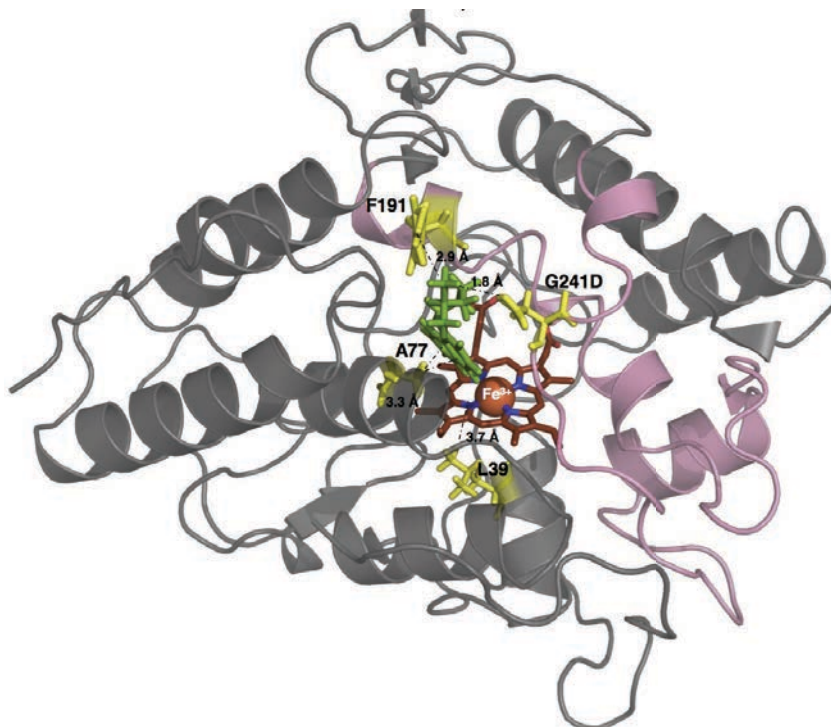

Figure 1. Selected region for focused evolution. The UPO structure (JaWa mutant) is shown as a grey cartoon and the relevant amino acids are indicated in yellow, together with the distances between them and the propranolol molecule (in green) or to the heme group (in dark red). The F191 position and G241D substitution seem to be involved in positioning the aliphatic branch of propranolol, while A77 apparently interacts with the aromatic rings and L39 lies underneath the heme. The D187-V248 segment (in pink) was subjected to MORPHING, harboring the α-helix where F191 is located and the loop of G241D. Docking simulations were performed using Molecular Operating Environment (MOE) software[23] and the crystal structure of the evolved UPO at a resolution of 1.2 Å (provided by Prof. Julia Sanz, IQFR-CSIC, Madrid, unpublished material).

We constructed two mutant libraries with different mutational loads and screened them using the 4-AAP assay (**Figure S4 in the Supporting Information**). As selection criterion, we imposed that UPO variants had to oxidize propranolol exceeding a threshold of 1.5-fold of parental´s activity. After two consecutive re-screenings, the six selected clones that satisfied the activity threshold carried the same single substitution (F191S), which did not jeopardize thermostability as the T_{50} value for both the parental JaWa and the mutant was maintained at ~60ºC (**Figures S4 and S5 in the Supporting Information**). This mutation improved the formation of 5´-OHP in the absence of AA up to 230% and most importantly, the

ratio between both reactions enhanced 1.8-fold in the presence and absence of AA. Given that Phe191 and Phe76 are responsible for defining the entrance to the heme access channel, whereas the aromatic triad formed by Phe69, Phe121 and Phe199 is involved in orienting the substrate for catalysis, all these residues were studied by saturation mutagenesis using as template the F191S variant (**Figure S6 in the Supporting Information**).[25] We first designed a combinatorial saturation muta-genesis library of Phe191 and Phe76 in which the most active clones exclusively incorporated again the F191S substitution. When individual saturation mutagenesis was performed at Phe69, Phe121 and Phe199, a set of non-functional libraries was obtained (80% inactive clones), with no further beneficial substitutions identified **(Figure S7 in the Supporting Information)**
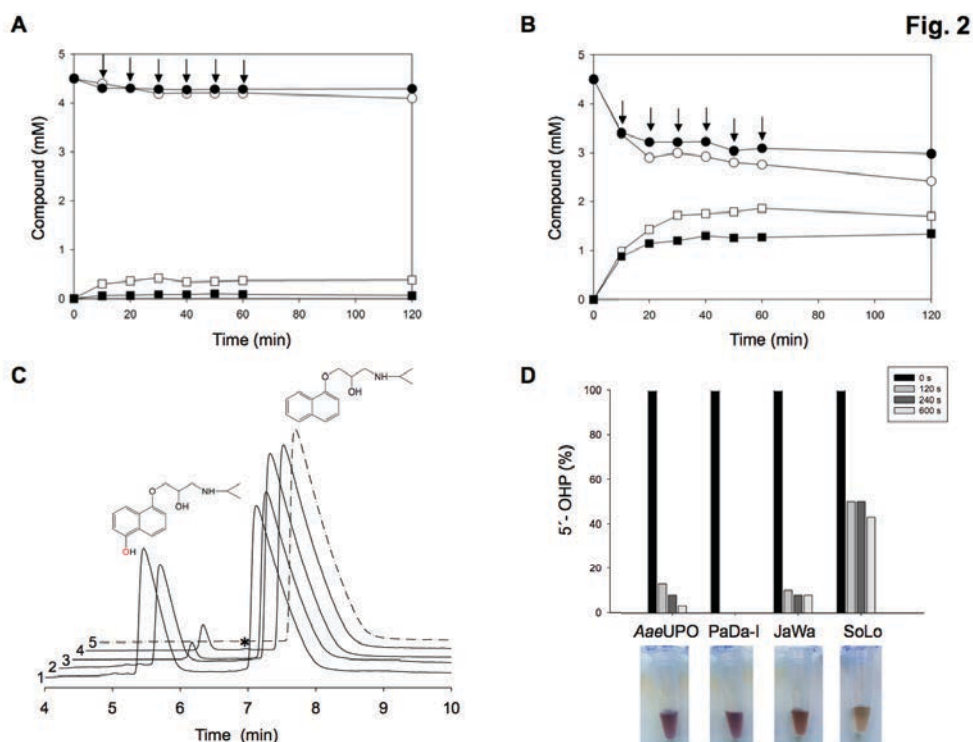
Figure 2. Propranolol conversion by native and mutant UPOs. (**A**) and (**B**) time course of the reactions over 120 min at pH 7.0 with the wildtype *Aae*UPO and the SoLo mutant, respectively. White circles, propranolol (with AA); black circles, propranolol (without AA); white squares, 5′-OHP (with AA); black squares, propranolol (without AA); arrows, periodic pulses of 2 mM H_2O_2. The total turnover numbers (TTN, μmol product/μmol en-zyme) were estimated from the 5′-OHP concentration after 120 min. (**C**) The HPLC elution profiles after a reaction time of 60 min in the absence of AA: 1, SoLo; 2, JaWa; 3, PaDa-I; 4, *Aae*UPO; 5, Control without enzyme; *, traces of DIP. (**D**) Conversion of 5′-OHP into

phenoxyl radicals and polymers. Reaction conditions can be found in the Materials and Methods section.

The mutant clone containing the F191S mutation (named SoLo), JaWa, PaDa-I and *Aae*UPO were all produced, purified to homogeneity [Reinheitszahl, *Rz* (A_{418}/A_{280})~2.2], and characterized biochemically and computationally at the atomic level. We first analyzed the transformation of propranolol by HPLC-PDA (**Figure 2c**) and notably, the regioselectivity for 5′-OHP shifted from 91% in *Aae*-UPO to 99% in PaDa-I, JaWa and SoLo. When boosting the UPO by periodic dosing with 2 mM H_2O_2 over the course of the reaction, without supplying AA, TTN of 3,000, 15,000 and 45,000 were achieved for the *Aae*UPO, JaWa and SoLo mutants, respectively, roughly representing a 15-fold improvement of SoLo relative to the wildtype *Aae*UPO (**Figure 2a, b**). More significantly, in the absence of AA the SoLo mutant still outperformed the TTN of the wildtype doped with AA more than 3-fold (*viz.* 45,000 and 14,000 TTN for SoLo in the absence of AA and *Aae*-UPO in the presence of AA, respectively).

Table 1. Steady kinetic parameters of UPO variants.

Substrate	Kinetic constant	*Aae*UPO	PaDa-I	JaWa	SoLo
ABTS	K_m (µM)	25 ± 2	50 ± 6	181 ± 22	568 ± 91
	k_{cat} (s^{-1})	221 ± 6	546 ± 19	125 ± 5	365 ± 23
	k_{cat}/K_m (M^{-1} s^{-1})	$8.8 \times 10^6 \pm 6.9 \times 10^5$	$11.0 \times 10^6 \pm 6.0 \times 10^5$	$6.9 \times 10^5 \pm 6.3 \times 10^4$	$6.4 \times 10^5 \pm 6.7 \times 10^4$
Propranolol	K_m (µM)	$2,239 \pm 333$	$2,268 \pm 220$	244 ± 92	391 ± 97
	k_{cat} (s^{-1})	150 ± 12	212 ± 11	765 ± 76	497 ± 35
	k_{cat}/K_m (M^{-1} s^{-1})	$6.7 \times 10^4 \pm 4.8 \times 10^3$	$9.3 \times 10^4 \pm 4.3 \times 10^3$	$3.1 \times 10^6 \pm 0.9 \times 10^5$	$1.3 \times 10^6 \pm 0.2 \times 10^5$
H_2O_2	K_m (µM)	$1,370 \pm 162$	$1,530 \pm 80$	$1,250 \pm 300$	$1,430 \pm 153$
	k_{cat} (s^{-1})	290 ± 15	676 ± 24	447 ± 40	446 ± 23
	k_{cat}/K_m (M^{-1} s^{-1})	$2.1 \times 10^5 \pm 1.5 \times 10^4$	$4.4 \times 10^5 \pm 12.0 \times 10^4$	$3.6 \times 10^5 \pm 5.9 \times 10^4$	$3.1 \times 10^5 \pm 1.8 \times 10^4$

Kinetic parameters were measured for the one-electron oxidation of ABTS (peroxidative activity), the two-electron oxidation of propranolol into 5′-OHP (peroxygenative activity) and for H_2O_2 with benzyl alcohol as peroxygenative substrate (**Table 1**). SoLo showed a striking 14-fold and 17-fold drop in the catalytic efficiency (k_{cat}/K_m) with ABTS relative to *Aae*UPO and PaDa-I, respectively, while maintaining similar performance as JaWa. The strong enhancement in the K_m, with a maximal 23-fold increase relative to *Aae*UPO was mostly responsible for this effect. Kinetic differences for propranolol were even more dramatic and the catalytic efficiencies for JaWa and SoLo were two orders of magnitude higher than for *Aae*UPO and PaDa-I, mostly due to the important 5- to 10-fold decrease in the K_m

for this substrate. The kinetics for H_2O_2 in the peroxygenation of benzyl alcohol did not differ among the UPO variants.
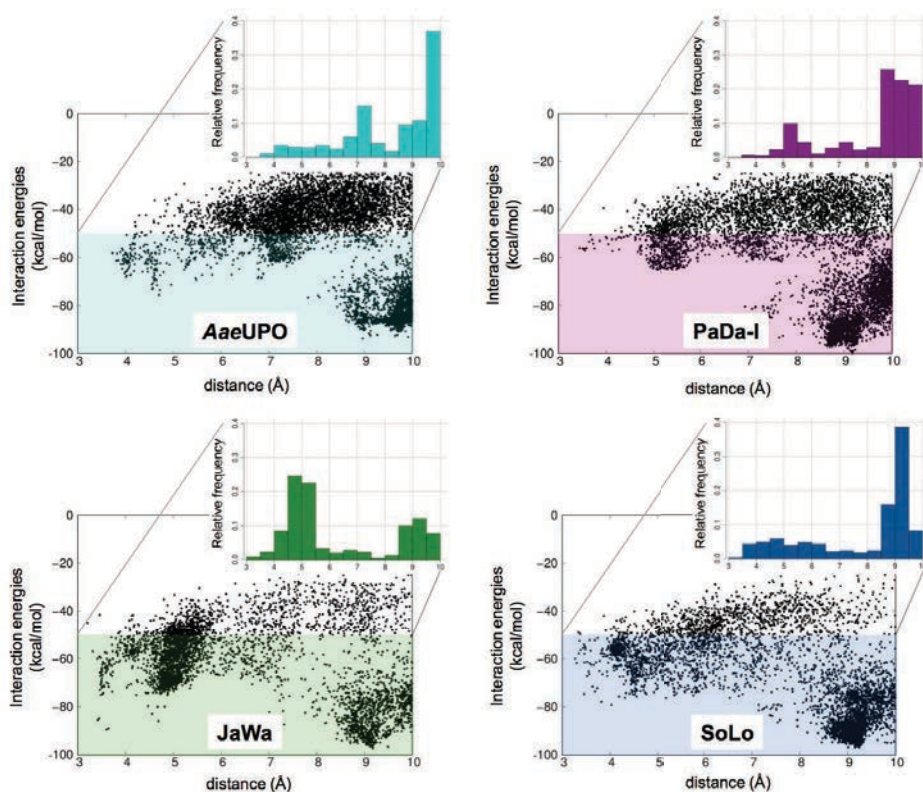

Figure 3. Computational analysis of propranolol diffusion in native and mutant UPOs. Interaction energies (in kcal·mol^{-1}) *vs.* ligand distances (in Å) from PELE simulations with propranolol in *Aae*UPO, PaDa-I, JaWa and SoLo UPO variants. The distances are between the reactive O atom in the heme compound I and the C_5 atom of propranolol. The top-right (inset) plots show the relative population for all the structures with interaction energies below -50 kcal·mol^{-1} using binning widths of 0.5 Å.

In order to study the differences found between the *Aae*UPO, PaDa-I, JaWa and SoLo variants during propranolol hydroxylation at the atomic level, we turned to molecular modeling employing PELE (Protein Energy Landscape Exploration), a Monte Carlo algorithm capable of effectively sampling the protein-ligand con-formational space.[26] We first modeled propanolol diffusion to the active site for each variant, finding that binding energies and distances largely correlate with the experimental K_m (**Figure 3 and Supporting Movie**). To facilitate oxygenation, ideally the distances between the C_5 propranolol carbon atom and the catalytic heme-ferryl oxygen should be around (or below) 4 Å. JaWa, with the lowest K_m for propranolol, clearly presents the best "catalytic minimum", with a distance

around 3.5 Å and a binding energy at *c.a.* -70kcal/mol. On the other hand, *Aae*UPO and PaDa-I, with noticeably higher K_m, have markedly worse binding energy profiles, which would result in the necessity of adding more substrate to reach correct catalytic positions. Moreover, when we analyzed the relative distance distribution of substrate's C_5 atom in the active site (**Figure 3**), we obtained a similar tendency: the relative frequency of structures below 4 Å is significantly higher for JaWa and SoLo variants. In addition, simulations provided important insights into the influence of the G241D mutation present in both JaWa and SoLo but absent in *Aae*UPO and PaDa-I (**Supporting Movie**), which facilitates the anchoring of the substrate in a more favorable orientation for hydroxylation. Besides, the Phe191 residue seems to tighten the stabilization of this catalytic orientation in the JaWa variant which is in agreement with the slight kinetic differences between both variants for propranolol. To obtain further structural insights into the changes caused by F191S mutation, 100 ns molecular dynamics (MD) were performed with JaWa and SoLo. MD pocket method[27] was used to track heme's cavity volume changes along the MD trajectories, showing a widening of the SoLo variant pocket compared to Jawa of ~50 Å^3 (**Figures 4C, D, E**). The most obvious reason behind the cavity broadening is the mutation of Phe191 to a less bulky amino acid such as serine. Moreover, structure visual inspection shows in JaWa, Phe191 tends to be placed in the hydrophobic heme cavity, causing a displacement of the α-helix hosting it and reducing the heme pocket volume. In the SoLo variant, such movement does not occur since Ser191 is kept buried into the protein by a hydrogen bond interaction within its hosting α-helix (**Figures 4A, B**).

However, we still wanted to ascertain why there was more 5´-OHP formed by SoLo than by JaWa, particularly given that the latter had an even higher catalytic efficiency for propranolol (**Table 1**). We hypothesized that the F191S mutation, which widens the access channel (**Figure 4**), could play an important role in bypassing the unwanted peroxidative activity on 5´-OHP (**2**) and the ensuing formation of the phenoxyl radical (**3**), thereby increasing the *P:p* ratio (numbering according to **Scheme 1**). To unveil the weaker peroxidative activity of SoLo, we measured the disappearance of **2** by HPLC-PDA during the reaction with the different variants in the absence of AA (**4**), using **2** as the departure reducing substrate. Pleasingly, the rate of converting **2** into **3** was noteworthy for all the variants except SoLo, which maintained a concentration of the substrate 10-fold higher than that of the rest of the variants after 10 minutes of the reaction (**Figure 2d**). These noticeable differences became readily visible because the polymeric products formed through the non-enzymatic coupling of the phenoxyl radicals are colored (see inset in **Figure 2d**). These results were confirmed computationally by PELE whereby the 5´-OHP diffusion addressed the significantly smaller concentration of this

peroxidative substrate at lower distances (<5 Å) from the heme catalytic center in SoLo compared to JaWa, as well as shorter residence times in the binding site (**Figure S8 in the Supporting Information**) which results in the inefficient 5´-OHP oxidation by SoLo variant. Taken together, our results unequivocally show that while the F191S mutation acts as the main driver of the remarkable decrease in peroxidative activity on **2**, the G241D change is mostly responsible for the improved peroxygenative activity on **1 (Figure S9 in the Supporting Information).**
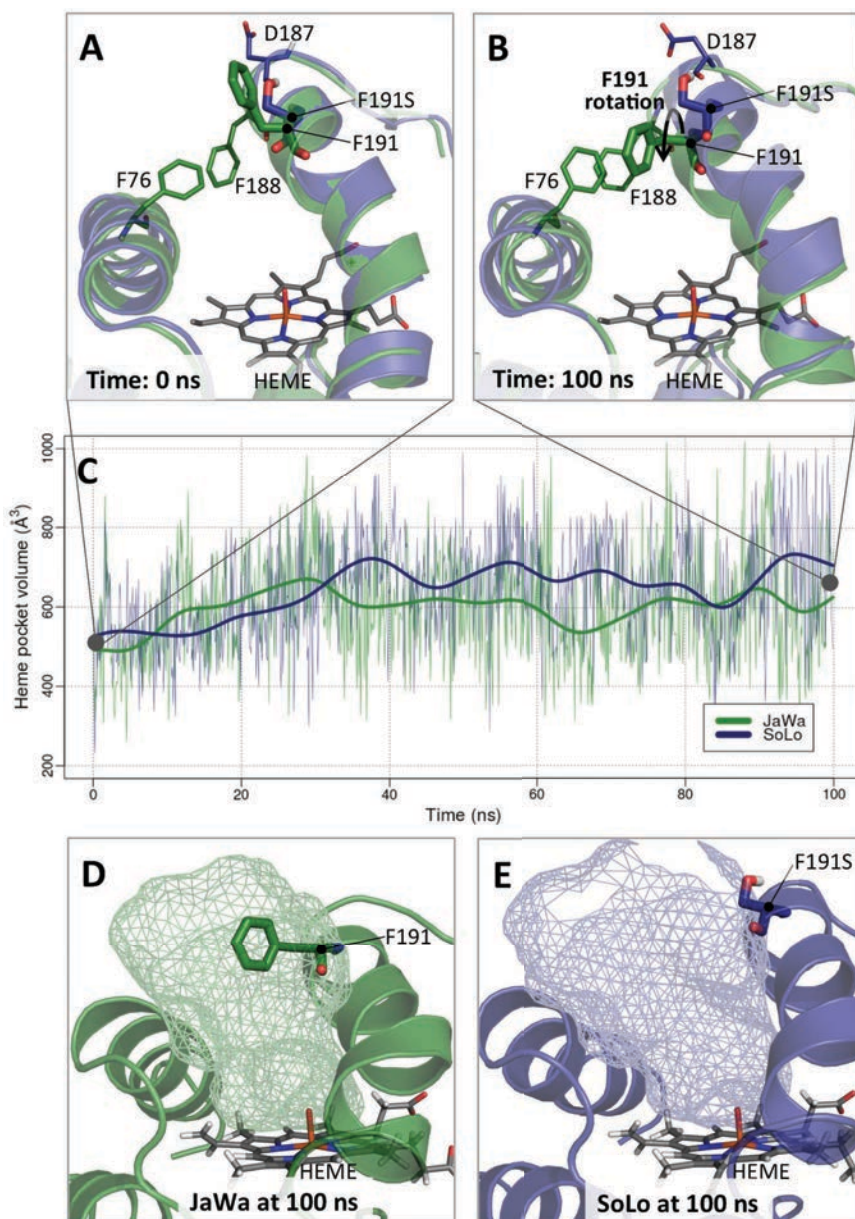
Figure 4. F191S induced structural changes along 100 ns MD. (A) and (**B**) JaWa and SoLo (colored in green and dark blue respectively) comparison at the beginning (0 ns, panel **A**) and at the end (100 ns, panel **B**) of the MD simulation. Simulations show that while F191 residue in JaWa causes a displacement of the α-helix hosting it and leads to the reduction of the heme cavity, in the SoLo variant F191S avoids the heme pocket hydrophobicity and tends to be buried into the α-helix holding it, interacting mainly with nearby residues (an hydrogen bond with D187 is conserved 87% among the MD). (**C**) Heme cavity volume tracking along 100 ns MD simulations for JaWa and SoLo variants, calculated with MDpocket every 0.1 ns. Smoothed volume of the pocket is shown in thick lines. (**D**) and (**E**) Heme binding pocket volume representation as a mesh surface for JaWa (Panel **D**) and SoLo (panel **E**) variants after 100 ns MD simulation.

To address the well known lability of heme-containing enzymes (including UPOs) against H_2O_2, we evaluated two strategies using either *in situ* generation of H_2O_2 from ambient O_2[28] or using *tert*-butyl hydroperoxide (*tert*-BuOOH) as milder peroxide source.[29]

For *in situ* H_2O_2 generation we utilized a bienzymatic cascade comprising alcohol oxidase (AOx) and formaldehyde dismutase (FDM) for the double oxidation of methanol to formic acid (generating two equivalents of H_2O_2, **Scheme 2**).[28] With such system, TTN as high as 264,000 and 226,000 were achieved with and without AA, respectively. Also the second strategy, *i.e.* using *tert*-BuOOH as milder oxidant, proved successful (**Figure S10 in the Supporting Information**). On semi-reparative scale, 20 and 10 mM of 5´-OHP were produced (20 and 40% analytical yield, respectively) with and without AA, respectively. The final product was isolated and further purified via a one-step flash chromatography step yielding 10.5 mg of pure 5´-OHP (as confirmed by ^{1}H-NMR, **Figure S11 in the Supporting Information**) in overall 15.2 % isolated yield. It should, however, be emphasized that neither the synthetic reaction, nor the isolation and purification procedure were optimized.

CONCLUSIONS AND OUTLOOK

Using H_2O_2 as a final electron acceptor and exclusive oxygen donor, UPO carries out a variety of oxygen-transfer reactions and as such, it is considered by many to be "taking the baton" from P450s in the field of synthetic organic chemistry.[19,29-31] In this study, we designed a highly active and stable UPO variant that behaves as a self-sufficient and efficient biocatalyst for the selective synthesis of 5´-OHP, irrespective of the presence of radical scavengers in the reaction. The UPO mutant shows the highest regioselectivity and TTN for the synthesis of 5´-OHP reported so far. When we compare it with the best engineered P450 BM3 heme

domain peroxygenase, SoLo mutant surpasses it roughly by 9,000-fold.[10-14] More significantly, with the assistance of a H_2O_2 *in situ* generation system based on an enzyme cascade reaction with methanol as the sacrificial electron donor for the reductive activation of O_2,[28] this UPO variant outperforms any natural or engineered hydroxylating catalyst described to date going one step closer to the industrial needs.[32]

MATERIAL & METHODS

Materials

ABTS, benzyl alcohol, Taq DNA polymerase, ascorbic acid and the Yeast Transformation kit were purchased from Sigma-Aldrich/Merck (Darmstadt, Germany). 5′-hydroxypropranolol (5′-OHP) was purchased from Santa Cruz Biotechnology (Santa Cruz, CA, USA) and DL-propranolol hydrochloride was purchased from ACROS Organics (Belgium). The high-fidelity DNA polymerase PfuUltra was acquired from Agilent. The BamHI and XhoI restriction enzymes were purchased from New England Biolabs (MA, USA) and the protease-deficient *S. cerevisiae* strain BJ5465 from LGCPromochem (Barcelona, Spain). The Zymoprep Yeast Plasmid Miniprep kit and Zymoclean Gel DNA Recovery kit were from Zymo Research (Orange, CA, USA). The NucleoSpin Plasmid kit was purchased from Macherey–Nagel (Düren, Germany) and the oligonucleotides were synthesized by Metabion (Bayern, Germany). All chemicals were reagent-grade purity or analytical standards.

Culture Media

The selective expression medium (SEM) contained filtered yeast nitrogen base (100 mL, 6.7%), filtered yeast synthetic drop-out medium supplement without uracil (100 mL, 19.2 g/L), filtered KH_2PO_4 buffer (pH 6.0, 67 mL, 1 M), filtered galactose (111 mL, 20%), filtered $MgSO_4$ (22 mL, 0.1 M), absolute ethanol (31.6 mL), filtered chloramphenicol (1 mL, 25 g/L), and *dd*H_2O (to 1000 mL). Sterile minimal medium contained 100 mL 6.7% filtered yeast nitrogen base, 100 mL 19.2 g/L filtered yeast synthetic drop-out medium supplement without uracil, 100 mL filtered 20% raffinose, 700 mL *dd*H_2O and 1 mL 25 g/L filtered chloramphenicol. SC drop-out plates contained 100 mL 6.7% filtered yeast nitrogen base, 100 mL 19.2 g/L filtered yeast synthetic drop-out medium supplement without uracil, 20 g autoclaved bacto agar, 100 mL 20% filtered glucose, 1 mL 25 g/L filtered chloramphenicol and *dd*H_2O to 1,000 mL. Sterile expression medium contained 720 mL autoclaved YP, 67 mL 1 M filtered KH_2PO_4 pH 6.0 buffer, 111 mL 20% filtered galactose, 22 mL filtered $MgSO_4$ 0.1 M, 300 mg/L of filtered hemoglobin,

31.6 mL absolute ethanol, 1 mL 25 g/L filtered chloramphenicol and *dd*H$_2$O to 1,000 mL. YP medium contained 10 g yeast extract, 20 g peptone and *dd*H$_2$O to 650 mL. YPD solution contained 10 g yeast extract, 20 g peptone, 100 mL 20% sterile glucose, 1 mL 25 g/L chloramphenicol and *dd*H$_2$O to 1,000 mL. Luria-Bertani (LB) medium was prepared with 5 g yeast extract, 10 g peptone, 10 g NaCl, 100 mg ampicillin and *dd*H$_2$O to 1,000 mL.

Laboratory evolution

The parent JaWa was obtained as described elsewhere.[16] All PCR products were loaded onto a preparative agarose gel and purified by use of the Zymoclean Gel DNA Recovery kit. The recovered DNA fragments were cloned under the control of the GAL1 promoter of the pJRoC30 expression shuttle vector, with use of BamHI and XhoI to linearize the plasmid and to remove the parent gene. The linearized vector was loaded onto a preparative agarose gel and purified with the Zymoclean Gel DNA Recovery kit.

Focused evolution at D187-V248 segment: The region between Asp187-Val248 selected from docking experiments was subjected to MORPHING (<u>m</u>utagenic <u>or</u>ganized <u>r</u>ecombination <u>p</u>rocess by <u>h</u>omologous *in vivo* <u>g</u>rouping).[24] Two Mutagenic PCR were prepared in a final volume of 50 µL containing 3% DMSO, 90 nM MJaWa Fw (5'-gcgcattcaagactccattg-3'), 90 nM MJaWa Rev (5'-gatcttgccgacatttttcc-3'), 0.3 mM dNTPs (0.075 mM each), MnCl$_2$ (mutational loads of 0.1 mM and 0.2 mM), 1.5 mM MgCl$_2$, 0.05 U/µL Taq DNA polymerase, and 1 ng/µl template. The amplification parameters were 94°C for 2 min (1 cycle); 94°C for 45 s, 48°C for 30 s, and 72°C for 90 s (28 cycles); and 72°C for 10 min (1 cycle). The remaining portions of the whole JaWa gene were amplified by high-fidelity PCR in a final volume of 50 µL containing 3% DMSO, 0.5 µM HFJaWa Fw (5'-caggctcatcctatgcagccc-3') and 0.5 µM RMLC (5'-gggagggcgtgaatgtaagc-3') or 0.5 µM HFJaWa Rev (5'-caaaggagaaattggggttggtcg-3') and 0.5 µM RMLN (5'-cctctatactttaacgtcaagg-3') for the other high fidelity fragment, 1 mM dNTPs (0.25 mM each), 0.05 U/µL PfuUltra DNA polymerase, and 2 ng/µL template. High-fidelity PCR was carried out on a gradient thermocycler using the following parameters: 95°C for 2 min (1 cycle); 95°C for 45 s, 48°C for 30 s, and 72°C for 90 s (28 cycles); and 72°C for 10 min (1 cycle). The whole gene was in vivo reassembled and recombined by transforming the different PCR products into *S. cerevisiae* competent cells, a process facilitated by ~40 bp overhangs flanking each recombination area.[33] The DNA transformation mixture was composed of linearized plasmid (100 ng) mixed with the mutagenized fragment (200 ng) and both

nonmutagenized fragments (200 ng). 1220 individual clones were screened (610 clones per mutant library).

Combinatorial saturation mutagenesis (F76 and S191): three PCR reactions were carried out in a final volume of 50 μL containing 3% DMSO, 0.3 mM dNTPs (0.075 mM each), 0.05 U/μL PfuUltra DNA polimerase, and 2 ng/μL template but each of them with different primers. PCR 1 with 0.25 μM of RMLN, 0.25 μM of F76 VHG R (5'-gcaagtccgtaatgagattgccgtccacaaggtgggccgcatatgtggc*cdb*gattgcggc-3'), 0.25 μM of F76 NDT R (5'-<u>gcaagtccgtaatgagattgccgtccacaaggtgggccgc</u>atatgtggc*ahn*gattgcggc-3') and 0.25 μM of F76 TGG R (5'-<u>gcaagtccgtaatgagattgccgtccacaaggtgggccgc</u>atatgtggc*cca*gattgcggc-3'). PCR 2 with 0.25 μM of HF F (5'-<u>gcggcccaccttgtggacggcaatctcattacggacttgc</u>-3'), 0.25 μM of S191 VHG R (5'-<u>cccatccacaaaaagattcgcggggaaggtggtctcgccg</u>taagcagt*cdb*gaacctaaag-3'), 0.25 μM of S191 NDT R (5'-<u>cccatccacaaaaagattcgcggggaaggtggtctcgccg</u>taagcagt*ahn*gaacctaaag-3') and 0.25 μM of S191 TGG R (5'-<u>cccatccacaaaaagattcgcggggaaggtggtctcgccg</u>taagcagt*cca*gaacctaaag-3'). PCR 3 with 0.25 μM of HF F-RMLC (5'-<u>cggcgagaccaccttccccgcgaatctttttgtggatggg</u>-3') and 0.25 μM of RMLC. Codon substitutions are shown in italics (where N = A/T/C/G; D = no C; V = no T, H = no G; y B = no A).[34] PCR reactions were carried out on a gradient thermocycler using the following parameters: 95°C for 2 min (1 cycle); 95°C for 45 s, 48°C for 45 s, and 72°C for 60 s (28 cycles); and 72°C for 10 min (1 cycle). Combinatorial saturation mutagenesis library was transformed into *S. cerevisiae* and the whole genes were *in vivo* reassembled and recombined by transforming the different PCR products into *S. cerevisiae* competent cells, a process facilitated by ~40 bp overhangs (underlined in each primer) flanking each recombination area.[33] The DNA transformation mixtures were composed of linearized plasmid (100 ng) mixed with the rest of the appropriate fragments (200 ng). 1480 clones were screened according to 22-trick method guidelines.[34]

Saturation mutagenesis at F69, F121, F199: saturation mutagenesis was carried out using degenerated NNK codons (N = A/T/C/G; K = T/G, M = A/C), creating three different libraries. Library F69: two PCR reactions were carried out in a final volume of 50 μL containing 3% DMSO, 0.2 mM dNTPs (0.05 mM each), 0.02 U/μL iProof DNA polymerase, 2 ng/μL template (SoLo) with 0.5 μM RMLN and 0.5 μM F69 R (5'-gaagattgcggcttgattgtcmnnattgaatc-3'), or 0.5 μM RMLC and 0.5 μM F69 F (5'-cgcggttcaggaaggattcaatnnkgacaatc-3'). Library F121: two PCR reactions were carried out in a final volume of 50 μL containing 3% DMSO, 0.2 mM dNTPs (0.05 mM each), 0.02 U/μL iProof DNA polymerase, 2 ng/μL template (SoLo) with 0.5 μM RMLN and 0.5 μM F121 R (5'-catactggcgtcgccttcmnnggtgccatgc-3'), or 0.5 μM RMLC and 0.5 μM F121 F (5'-ggactcaatgagcatggcaccnnkgaaggcg-3').

Library F199: two PCR reactions were carried out in a final volume of 50 μL containing 3% DMSO, 0.2 mM dNTPs (0.05 mM each), 0.02 U/μL iProof DNA polymerase, 2 ng/μL template (SoLo) with 0.5 μM RMLN and 0.5 μM F199 R (5'-ccacaaaaagattcgcgggmnnggtggtctcg-3'), or 0.5 μM RMLC and 0.5 μM F199 F (5'-ctactgcttacggcgagaccaccnnkcccgcg-3'). PCR reactions were carried out on a gradient thermocycler using the following parameters: 98°C for 30 s (1 cycle); 98°C for 10 s, 48°C for 30 s, and 72°C for 30 s (28 cycles); and 72°C for 10 min (1 cycle). Each library was transformed independently and the whole genes were in vivo reassembled and recombined by transforming the different PCR products into *S. cerevisiae* competent cells, a process facilitated by ∼40 bp overhangs flanking each recombination area.[33] The DNA transformation mixtures were composed of linearized plasmid (100 ng) mixed with the rest of the appropriate fragments (200 ng). Each library contained 170 clones.

High Throughput Dual screening: Individual clones were picked and inoculated in sterile 96-well plates (Greiner Bio-One, GmbH, Germany), referred to as master plates, containing 200 mL of SEM per well. In each plate, column number 6 was inoculated with the parent type, and one well (H1-control) was inoculated with *S. cerevisiae* transformed with pJRoC30-MtL plasmid (laccase without activity). Plates were sealed with parafilm to prevent evaporation and incubated at 30°C, 220 RPM and 80% relative humidity in a humidity shaker (Minitron, Infors, Switzerland) for five days. The master plates were centrifuged (Eppendorf 5810R centrifuge, Germany) for 10 min at 2,500g and 4°C. Aliquots of the supernatants (20 μL) were transferred from the master plates to two replica plates with the aid of a liquid handler robotic station Freedom EVO (Tecan, Switzerland). 50 μL of the reaction mixture (with or without ascorbic acid) were added to the two replica plates with the help of a pipetting robot (Multidrop Combi Reagent Dispenser, Thermo Scientific, MA, USA). Reaction mixture for replica plate 1 contained 50 mM potassium phosphate buffer (pH 7.0), 5 mM propranolol and 2 mM H_2O_2. Reaction mixture for replica plate 2 contained 50 mM potassium phosphate buffer (pH 7.0), 5 mM propranolol, 2 mM H_2O_2 and 4 mM ascorbic acid. Replica plates 1 and 2 were incubated at room temperature for 30 and 60 min, respectively. Afterwards, the amount of 5'-OHP formed in each well was determined by the 4 aminoantipyrine (4-AAP) assay[22] (**Figures S1-S3 in the Supporting Information**). Plates were stirred briefly and absorption at 530 nm was recorded in a plate reader (SPECTRA-Max Plus 384, Molecular Devices, USA). The values were normalized against the parent type of the corresponding plate and selected variants came from the ratio between the values obtained in the absence (peroxygenative+peroxidative activities) and in the presence (peroxygenative activity) of ascorbic acid. To rule out the

selection of false positives, two re-screenings were carried out as described else-where.[20]

Production, purification and biochemical characterization

*Aae*UPO wildtype was produced and purified as described before.[11]

Production of recombinant UPO variants in *S. cerevisiae*: A single colony from the *S. cerevisiae* clone containing the parent or mutant gene was picked from a SC drop-out plate, inoculated in minimal medium (10 mL) and incubated for 48 h at 30°C and 230 RPM. An aliquot of cells was removed and used to inoculate minimal medium (100 mL) in a 500 mL (at a OD_{600} of 0.25). The cells completed two growth phases (6–8 h) and then expression medium (900 mL) was inoculated with the pre-culture (100 mL) (OD_{600} of 0.1). After incubating for 72 h at 25°C and 230 RPM (maximal UPO activity; $OD_{600} = 25$-30), the cells were recovered by centrifugation at 4,500 RPM (4°C) and the supernatant was double-filtered (using both glass membrane and a nitrocellulose membrane of 0.45 µm pore size).

Purification of recombinant UPO variants: Recombinant UPO purification was achieved by cationic exchange chromatography and anion exchange chromatography (ÄKTA purifier, GE Healthcare, WI, US). The crude extract was concentrated and dialyzed in sodium phosphate/citrate 20 mM at pH 3.3 (buffer A) by tangential ultrafiltration (Pellicon; Millipore, Temecula, CA, US) through a 10-kDa-pore-size membrane (Millipore) by means of a peristaltic pump (Masterflex Easy Load; Cole-Parmer, Vernon Hills, IL). The sample was filtered and loaded onto a strong cation-exchange column (HiTrap SP FF GE Healthcare) pre-equilibrated with buffer A. The proteins were eluted with a linear gradient from 0 to 40% of buffer A within 60 mL of NaCl and from 40 to 100% within 5 mL at a flow rate of 1 mL/min. Fractions with UPO activity *vs.* ABTS were harvested, concentrated, dialyzed against buffer Tris HCl 20 mM at pH 7.8 (buffer B) and loaded onto a 10 µm high resolution anion-exchange Biosuite Q column (Waters) pre-equilibrated with buffer B. The proteins were eluted with a linear gradient from 0 to 20% within 40 mL of NaCl and from 20 to 100% within 5 mL at a flow rate of 1 mL/min. The fractions with UPO activity *vs.* ABTS were pooled, dialyzed against buffer potassium phosphate 10 mM at pH 7.0 and stored at 4°C.

Steady-state kinetic constants: ABTS kinetic constants for UPO were estimated in sodium phosphate/citrate buffer (pH 4.0, 100 mM), containing H_2O_2 (2 mM), and for propranolol in potassium phosphate buffer (pH 7.0, 50 mM) containing H_2O_2 (2 mM). For H_2O_2, benzyl alcohol was used as a reducing substrate under the

corresponding saturated conditions. Reactions were performed in triplicate, and substrate oxidations were followed through spectrophotometric changes (ε_{418}ABTS$^{•+}$=36,000 M^{-1} cm^{-1}; ε_{325} 5´-OHP=1,996 M^{-1} cm^{-1} and ε_{280} benzalde-hyde=1,400 M^{-1} cm^{-1}). To calculate the K_m and k_{cat} values, the average V_{max} was represented against substrate concentration and fitted to a single rectangular hyper-bola function with SigmaPlot 10.0, where parameter a was equal to k_{cat} and param-eter b was equal to K_m.

HPLC analysis: The reaction mixtures were analyzed by reversed-phase chroma-tography (HPLC) with equipment consisting of a tertiary pump (Varian/Agilent Technologies) coupled to an autosampler (Merck Millipore) and an Zorbax Eclipse plus C18 (15 cm x 4,6 cm) column at 40°C. Detection was performed with a PDA (Varian/Agilent Technologies). The mobile phase was a gradient from 10% meth-anol and 90% ddH$_2$O (both with 0.1% acetic acid) to 90% methanol and 10% ddH$_2$O at a flow rate of 0.8 mL/min. The reaction was quantified at 280 nm (from HPLC standards). Reaction mixtures containing 0.03 µM purified enzyme, 5 mM propranolol, and 2 mM H$_2$O$_2$ in 50 mM potassium phosphate pH 7.0 (0.5 mL final volume). After 60 minutes at room temperature, reaction was stopped by heating 10 min at 70°C and cooling 5 min at 4°C (see **Figure 2C**). Total turnover numbers were calculated from reaction that contained 0.03 µM purified enzyme, 5 mM pro-pranolol, and pulses of 2 mM H$_2$O$_2$ every 10 min in 50 mM potassium phosphate buffer pH 7.0 (0.5 mL final volume). Reactions were stopped at different times (10 to 120 min) as described before (see **Figure 2A, B**). In the determination of 5´-hydroxypropranolol (5´-OHP) consumption by UPO, the mixture contained 0.03 µM purified enzyme, 0.5 mM 5´-OHP, 20% methanol 20% and 2 mM H$_2$O$_2$ in 50 mM potassium phosphate buffer pH 7.0 (0.5 mL final volume). The reaction was started by the addition of the H$_2$O$_2$ and stopped at different times (2 to 10 min) as described before (see **Figure 2D**). Samples of each experiment (20 µL) were in-jected and analyzed. Standard deviations were lower than 5% in all experiments. Products were identified and quantified against authentic standards by the HPLC method abovementioned.

Kinetic thermostability: The thermostability of the different UPO samples was es-timated by assessing their T_{50} values using 96/384 well gradient thermocyclers. Appropriate UPO dilutions were prepared with the help of the robot in such a way that 20 µL aliquots gave rise to a linear response in the kinetic mode. Then, 50 µL were used for each point in the gradient scale and a temperature gradient profile ranging from 30 to 80°C was established as follows (in °C): 30.0, 31.7, 34.8, 39.3, 45.3, 49.9, 53.0, 55.0, 56.8, 59.9, 64.3, 70.3, 75.0, 78.1 and 80.0. After a 10 min incubation, samples were chilled out on ice for 10 min and further incubated at

room temperature for 5 min. Afterwards, 20 µL of samples were subjected to the ABTS colorimetric assay described elsewhere.[20] The thermostability values were deduced from the ratio between the residual activities incubated at different temperature points and the initial activity at room temperature.

Reaction optimization and product identification

Production of 5´-OHP by SoLo coupled to an *in situ* H_2O_2 generation system: For *in situ* H_2O_2 generation, 10 nM of SoLo mutant, 10 nM alcohol oxidase from *Pichia pastoris* (AOx) and 600 nM formate dismutase from *Pseudomonas putida* (FDM) in combination with 200 mM methanol were used.[28] Reactions were performed at 30°C and 600 rpm in 100 mM potassium phosphate buffer pH 7.0 with 10 mM of propranolol in a total volume of 375 µL using a thermo shaker device (Eppendorf, Germany). After 3 h, the reactions were stopped by heat inactivation for 10 min at 70°C, followed by cooling down the sample and separation of precipitate by centrifugation at 5°C. Quantification of 5´-OHP was carried out by HPLC analytics. HPLC measurements were performed at 40°C on a Shimadzu LC-20 system with a Shimadzu SPD-M20A Photo Diode Array detector using Waters Xterra RP18 column (4.6×150 mm, 3.5 µm). For the mobile phase acetonitrile (ACN) and water containing 5% ACN and 0.1% trifluoroacetic acid (TFA) was used. The separation was performed in an isocratic mode at a flow rate of 1.0 mL/min while increasing the ACN concentration in 3 steps to finally 100%: 25%, 4 min hold, 50% 4 min hold and 100%, 3 min hold. The reaction was quantified at 280 nm based on a calibration using 5´-OHP standard.

Semi-preparative production of 5´-OHP: For the semi-preparative production 5´-OHP, 50 mM propranolol was dissolved in 100 mM potassium phosphate buffer pH 7.0 (10 mL total volume). Reactions were performed at 30°C while gently mixing (neoLab rotator, mode C2, 20 rpm, Germany) in presence or absence of 40 mM of AA using 0.5 µM of SoLo. The reaction was started by the addition of 5 mM *tert*-butyl hydroperoxide (*tert*-BuOOH). Every hour, 5 mM *tert*-BuOOH was added and 200 µL of each reaction time were taken for 5´-OHP quantification by HPLC. After 9 h, 5 mL of the reaction mixture were used for 5´-OHP isolation. The solution was acidified (200 µL, 37% HCl), extracted with ethyl acetate (3x 5 mL), dried with $MgSO_4$, and the solvent was removed under reduced pressure. The crude product was purified by column chromatography using silica gel (60 Å, 230-400 mesh) with a dichloromethane (DCM):methanol eluent (8:2). Separation was followed using thin layer chromatography (TLC) on silica gel-coated plates (Macherey-Nagel, Polygram SIL G/UV254) with a DCM:methanol solvent mixture (9:1).

NMR analytics: NMR spectra were recorded on an Agilent 400 (400 MHz) spectrometer in methanol-*d4*. Chemical shifts are given in ppm with respect to tetramethylsilane. Coupling constants are reported as J-values in Hz (**Figure S11 in the Supporting Information**).

Computational analysis

System preparation for molecular modeling: The starting structure for PELE (Protein Energy Landscape Exploration) simulations with *Aae*UPO was the *Agrocybe aegerita* peroxygenase crystal structure (PDB ID: 2YOR). For the different UPO mutants, the crystal structure of PaDa-I (at a resolution of 1.5 Å, courtesy of Prof. Julia Sanz, IQFR-CSIC, Madrid, unpublished material) was used for modeling. Since the optimal pH for propranolol peroxygenation and subsequent 5´-OHP peroxidation is 7.0, all the structures were prepared accordingly with the aid of the Schrödinger Protein Preparation Wizard[35] and the H++ web server.[36] All acidic residues were deprotonated. Histidines were δ-protonated, with the exception of His82 (ε-protonated) and His118 and His251 (double-protonated). To relax the systems after mutations insertion, and to investigate their possible effect on the protein structure, 5 ns of Molecular Dynamics simulation (MD) were performed with Desmond[37] on JaWa and SoLo structures. Finally, the heme site was modeled as compound I after being fully optimized in the protein environment with quantum mechanics/molecular mechanics (QM/MM) using QSite.[38] Propranolol and 5´-OHP molecules were also optimized with Jaguar[39] at the DFT/M06 level with the 6-31G** basis and PBF implicit solvent in order to obtain their electrostatic potential atomic charges.

Protein energy landscape exploration (PELE) computational analysis: Once all protein and ligand structures were prepared, heme binding site explorations were performed with PELE.[26] Substrates were manually placed in identical positions at the entrance of the heme-access channel, and PELE simulations were carried out in two different stages. First, ligands were driven from the solvent to the UPO heme binding site. Then, once the center of mass of the ligand was within 6 Å of the heme catalytic oxygen, it was free to explore the active site. The results presented here are based on 160 independent 48 h PELE simulations. Moreover, in an attempt to increase the sampling for 5´-OHP diffusion, additional PELE simulations were performed for this substrate. Representative structures at distances lower than 10 Å (considering the distance between the reactive O-heme atom and the H_5 substrate atom) were selected using binning widths of 1 Å from previous PELE simulations and used as starting structures for the new ones. New simulations were setup to

freely explore the active site and a total of 196 independent 48 h simulations were run.

Molecular Dynamics and MDpocket: To study the changes caused by F191S mutation, 100 ns MD simulations were performed with JaWa and SoLo using Desmond.[37] To prepare the system for MD, previously mentioned prepared JaWa and SoLo systems were placed inside an orthorhombic box containing SPC explicit waters and ions to neutralize the system at a concentration of 0.15 M NaCl. From the solvated system, MD simulations were run with the following parameters: the OPLS-2005 force field, the temperature was regulated with a Nose-Hoover chain thermostat with a relaxation time of 1.0 ps, the pressure with the Martyna-Tobias-Klein barostat with isotropic coupling and a relaxation time of 2.0 ps, and finally, the production phase was run over the course of 100 ns using the NPT canonical ensemble at 300K. From MD simulations, structures at every 0.1 ns were extracted and used for volume pocket calculation with MDpocket,[27] a fast and open-source tool for protein pocket (cavity) detection on molecular dynamic trajectories or other conformational ensembles.

ASSOCIATED CONTENT

Supporting Information

Supporting Figures S1-S11; Supporting Movie.

ACKNOWLEDGEMENTS

COMPETING FINANCIAL INTERESTS

The authors have no competing financial interests to declare.

REFERENCES

1. National Research Council. *Industrialization of Biology: A Roadmap to Accelerate the Advanced Manufacturing of Chemicals*, The National Academies Press: Washington DC, **2015**.

2. Baillie, T. A.; Cayen, M. N.; Fouda, H.; Gerson, R. J.; Green, J. D.; Grossman, S. J.; Klunk, L. J.; LeBlanc, B.; Perkins, D. G.; Shipley, L. A. Drug metabolites in safety testing. *Toxicol. Appl. Pharmacol*. **2002**, 182, 188-196.

3. Atrakchi, A. H. Interpretation and considerations on the safety evaluation of human drug metabolites. *Chem. Res. Toxicol*. **2009**, 22, 1217-1220.

4. Guengerich, F. P. Safety Assessment of Stable Drug Metabolites. *Chem. Res. Toxicol*. **2009**, 22, 237-238.

5. Al-Majed, A. A.; Bakheit, A. H. H.; Abdel Aziz, H. A.; Alajmi, F. M.; AlRabiah, H. In: *Profiles of Drug Substances, Excipients and Related Methodology* Brittain H. G. Ed;. Elsevier, United Kingdom, **2017**, 42, 287-338.

6. Greenslade, F. C.; Newquist, K. L. In vitro measurement of the beta-adrenergic blocking properties of ORF 12592, the 5-hydroxy analog of propranolol. *Arch. Int. Pharmacodyn. Ther*. **1978**, 233, 270-280.

7. Oatis, J. E. Jr; Russell, M. P.; Knapp, D. R.; Walle, T. Ring-Hydroxylated Propranolol: Synthesis and P-Receptor Antagonist and Vasodilating Activities of the Seven Isomers *J. Med. Chem*. **1981**, *24*, 309-314.

8. Kumamoto, T.; Aoyama, N.; Nakano, S.; Ishikawa, T.; Narimatsu, S. Synthesis of enantiomeric 4-hydroxypropranolols from 1,4-dihydroxynaphthalene. *Tetrahedron Asymmetry*, **2001**, 12, 791-795.

9. Eiben, S.; Kaysser, L.; Maurer, S.; Kühnel, K.; Urlacher, V. B.; Schmid, R. D. Preparative use of isolated CYP102 monooxygenases -- a critical appraisal. *J. Biotechnol*. **2006**, 124, 662-669.

10. Otey, C. R.; Bandara, G.; Lalonde, J.; Takahashi, K.; Arnold, F.H. Preparation of human metabolites of propranolol using laboratory-evolved bacterial cytochromes P450. *Biotechnol. Bioeng*. **2006**, 93, 494-499.

11. Ullrich, R.; Nüske, J.; Scheibner, K.; Spantzel, J.; Hofrichter, M. Novel haloperoxidase from the agaric basidiomycete Agrocybe aegerita oxidizes aryl alcohols and aldehydes. *Appl. Environ. Microbiol*. **2004**, 70, 4575-4581.

12. Hofrichter, M.; Ullrich, R. Oxidations catalyzed by fungal peroxygenases. *Curr. Opin. Chem.* **2014**, 19, 116-125.

13. Poraj-Kobielska, M.; Kinne, M.; Ullrich, R.; Scheibner, K.; Kayser, G.; Hammel, K.E.; Hofrichter, M. Preparation of human drug metabolites using fungal peroxygenases. *Biochem. Pharmacol*. **2011**, 82*, 789-796.

14. Kinne, M.; Poraj-Kobielska, M.; Aranda, E.; Ullrich, R.; Hammel, K. E.; Scheibner, K.; Hofrichter, M. Regioselective preparation of 5-hydroxypropranolol and 4'-hydroxydiclofenac with a fungal peroxygenase. *Bioorg. Med. Chem. Lett*. **2009**, 19, 3085-3087.

15. Kiebist, J.; Holla, W.; Heidrich, J.; Poraj-Kobielska, M.; Sandvoss, M.; Simonis, R.; Gröbe, G.; Atzrodt, J.; Hofrichter, M.; Scheibner, K. One-pot synthesis of human metabolites of SAR548304 by fungal peroxygenases. *Bioorg. Med. Chem*. **2015**, 23, 4324-4332.

16. Molina-Espeja, P.; Cañellas, M.; Plou, F. J.; Hofrichter, M., Lucas, F.; Guallar, V.; Alcalde, M. Synthesis of 1-Naphthol by a Natural Peroxygenase Engineered by Directed Evolution. *Chembiochem* **2016**, 17, 341-349.

17. 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* **2017**, 30, 189-196.

18. Ullrich, R.; Hofrichter, M. Enzymatic hydroxylation of aromatic compounds. *Cell Mol. Life Sci.* **2007**, 64, 271-293.

19. Wang, Y.; Lan, D.; Durrani, R.; Hollmann, F. Peroxygenases en route to becoming dream catalysts. What are the opportunities and challenges? *Curr Opin Chem Biol* **2017**, 37, 1-9.

20. Molina-Espeja, P.; Garcia-Ruiz, E.; Gonzalez-Perez, D.; Ullrich, R.; Hofrichter, M.; Alcalde, M. Directed evolution of unspecific peroxygenase from Agrocybe aegerita. *Appl. Environ. Microbiol.* **2014**, 80, 3496-3507.

21. Molina-Espeja, P.; Ma, S.; Mate, D.M.; Ludwig, R.; Alcalde, M. Tandem-yeast expression system for engineering and producing unspecific peroxygenase. *Enz. Microb. Technol*. **2015**, 73-74, 29-33.

22. Otey, C. R.; Joern, J. M. High-throughput screen for aromatic hydroxylation. *Methods Mol. Biol.* **2003**, 230, 141-148.

23. Molecular Operating Environment (MOE); version 2010, Chemical Computing Group ULC, Canada, **2010**.

24. Gonzalez-Perez, D.; Molina-Espeja, P.; Garcia-Ruiz, E.; Alcalde M. Mutagenic Organized Recombination Process by Homologous IN vivo Grouping (MORPHING) for directed enzyme evolution. *PLoS ONE* **2014**, 9, e90919.

25. Piontek, K.; Strittmatter, E.; Ullrich, R.; Gröbe, G.; Pecyna, M. J.; Kluge, M.; Scheibner, K.; Hofrichter, M.; Plattner, D. A. Structural Basis of Substrate Conversion in a New Aromatic Peroxygenase. Cytochrome P450 functionality with benefits. *J. Biol. Chem.* **2013**, 288, 34767-34776.

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

27. Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics*, **2011**, 27, 3276-3285.

28. Ni, Y.; Fernandez-Fueyo, E.; Gomez-Baraibar, A.; Ullrich, R.; Hofrichter, M.; Yanase, H.; Alcalde, M.; van Berkel, W. J. H.; Hollman, F. Peroxygenase-Catalyzed Oxyfunctionalization Reactions Promoted by the Complete Oxidation of Methanol. *Angew. Chem. Int. Edit.* **2016**, 55, 798-801.

29. Bormann, S.; Gomez-Baraibar, A.; Ni, Y.; Holtmann, D.; Hollmann, F. Specific oxyfunctionalisations catalysed by peroxygenases: opportunities, challenges and solutions. *Catal. Sci. Technol.* **2015**, 5, 2038-2052.

30. Hofrichter, M.; Kellner, H.; Pecyna, M. J.; Ullrich R. In: *Monooxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome P450*; Hrycay, E. G. and Bandiera, S. M., Eds.; Advances in Experimental Medicine and Biology, USA. **2015**, Vol. 851, pp 341-368.

31. Molina-Espeja, P.; Gomez de Santos, P.; Alcalde M. In: *Directed Enzyme Evolution: Advances and Applications;* Alcalde, M., Ed.; Springer, Switzerland. **2017**, pp 127-143.

32. Gomez de Santos, P., Molina-Espeja, P., Plou, F.J. and Alcalde, M. Regioselective synthesis of 5´-hydroxypropranolol by aromatic peroxygenase designed by directed evolution. P300212672.

33. Alcalde, M. In: *In vitro Mutagenesis Protocols, 3rd ed. Methods in Molecular Biology*; Bramman, J., Ed; Springer-Humana Press: USA, **2010**, pp. 3-15.

34. Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z. G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2013**, 2, 83-92.

35. Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **2013**, 27, 221-234.

36. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acid Res.* **2012**, 40, W537-W541.

37. *Desmond Molecular Dynamics System,* version 2.2, D.E. Shaw Research, New York, NY, **2009**. Maestro-Desmond Interoperability Tools, version 2.2, Schrödinger, New York.

38. QSite, version 5.7, Schödinger, LLC, New York, **2011**.

39. Jaguar, version 8.1, Schrödinger, LLC, New York, **2013**.

Supporting Information

Selective synthesis of the human drug metabolite 5´-hydroxypropranolol by an evolved self-sufficient peroxygenase

Patricia Gomez de Santos[†], Marina Cañellas[‡], Florian Tieves[§], Sabry H. H. Younes[§], Patricia Molina-Espeja[†], Martin Hofrichter[φ], Frank Hollmann[§], Victor Guallar[‡] and Miguel Alcalde[†*]

[†]Department of Biocatalysis, Institute of Catalysis, CSIC, 28049 Madrid, Spain.

[‡]Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, 08034 Barcelona, Spain.

[§]Department of Biotechnology, Delft University of Technology, van der Massweg 9, 2629HZ Delft, The Netherlands.

[φ]Department of Bio- and Environmental Sciences, TU Dresden, International Institute Zittau, Mark 23, 02763 Zittau, Germany.

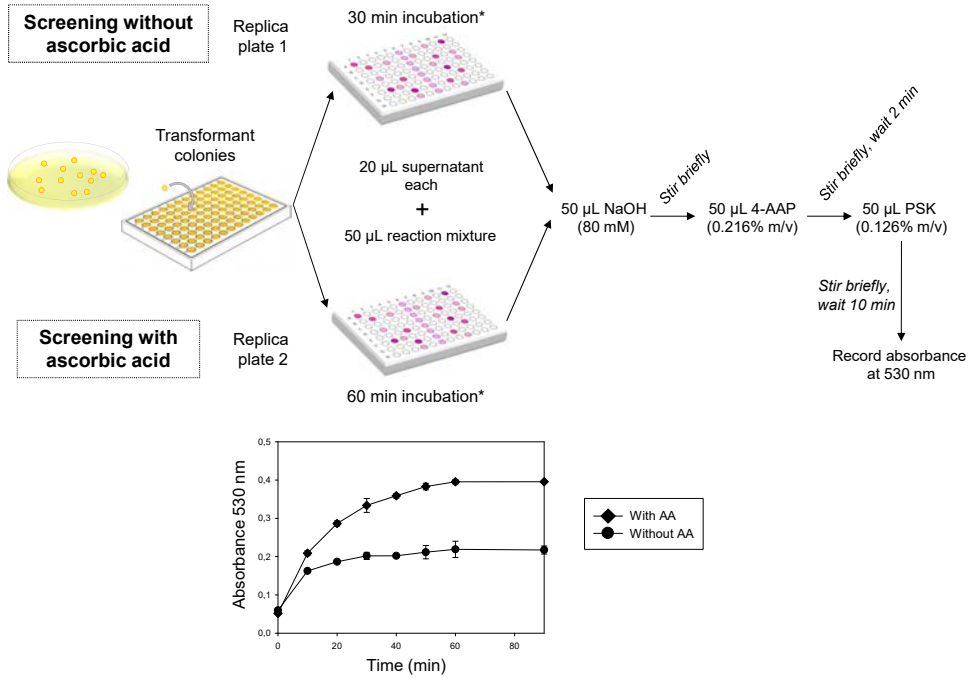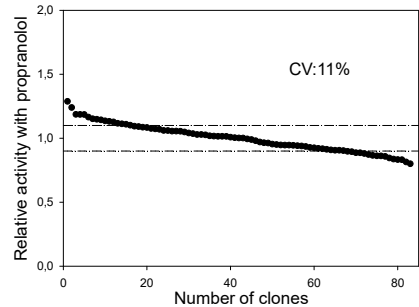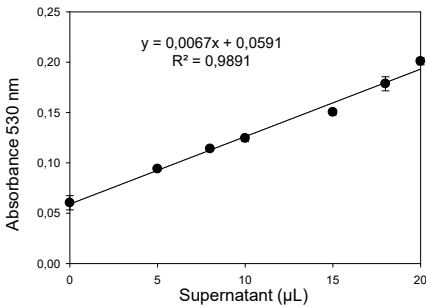* Correspondence should be addressed to: M.A. (malcalde@icp.csic.es).

Figure S1: Step-by-step screening protocol and optimization of the incubation time with and without AA.

Without ascorbic acid
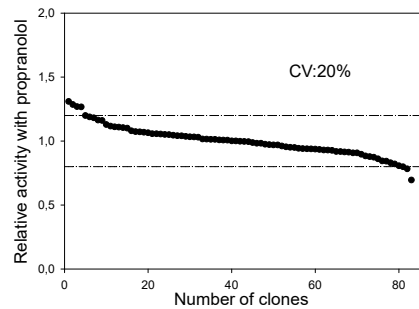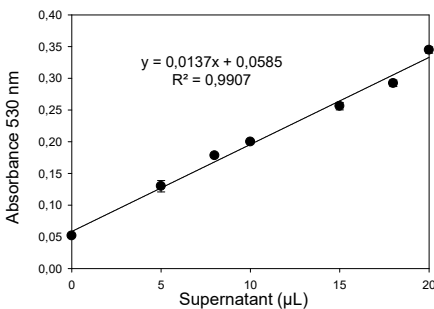

With ascorbic acid

Figure S2: Validation of the screening assay. (Left) Linearity of the assay, (Right) Coefficient of variance (CV) of the spectrophotometric assay. The landscapes correspond to 83 independent clones containing JaWa parental type, grown in microtiter format. The activity of the clones is plotted in descending order; dashed lines indicate the CV for each assay.
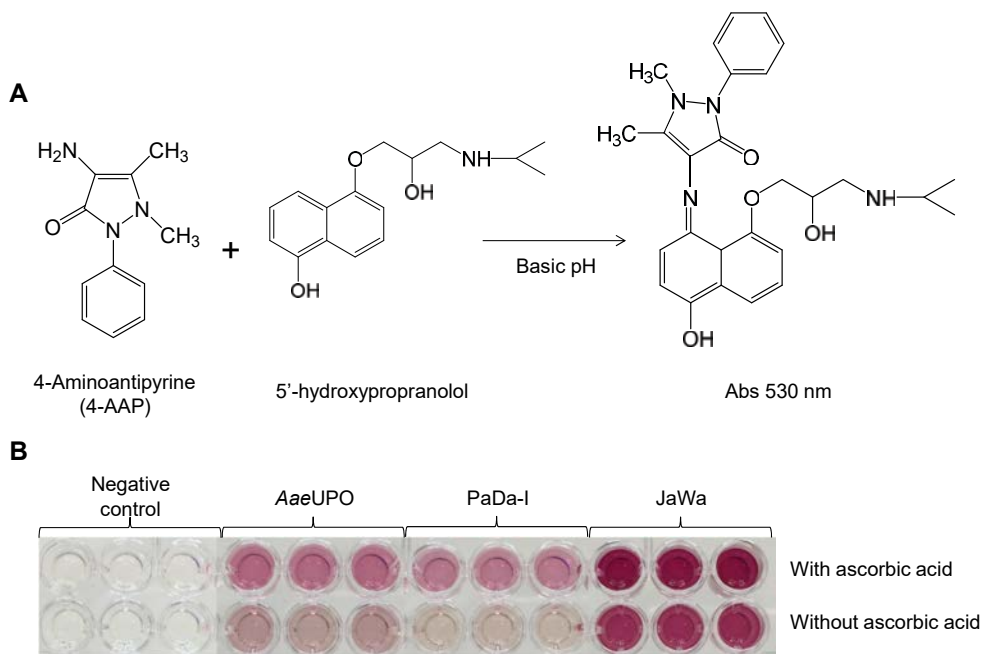
A

4-Aminoantipyrine
(4-AAP)

5'-hydroxypropranolol

Basic pH

Abs 530 nm

B

Negative control   *Aae*UPO   PaDa-I   JaWa

With ascorbic acid

Without ascorbic acid

Figure S3: Screening assay to detect 5´-OHP with 4-AAP. (**A**) The dual screening assay allowed us to select clones with weaker peroxidative activity on 5´-OHP and enhanced peroxygenase activity on propranolol. We used the ratio between the activities in the presence and absence of AA as a discriminatory factor of the assay. (**B**) Benchmarking of the reaction with propranolol and different UPO variants, in the presence and absence of AA.

A

B

Without AA   With AA

Control

JaWa parent

SoLo (F191S)

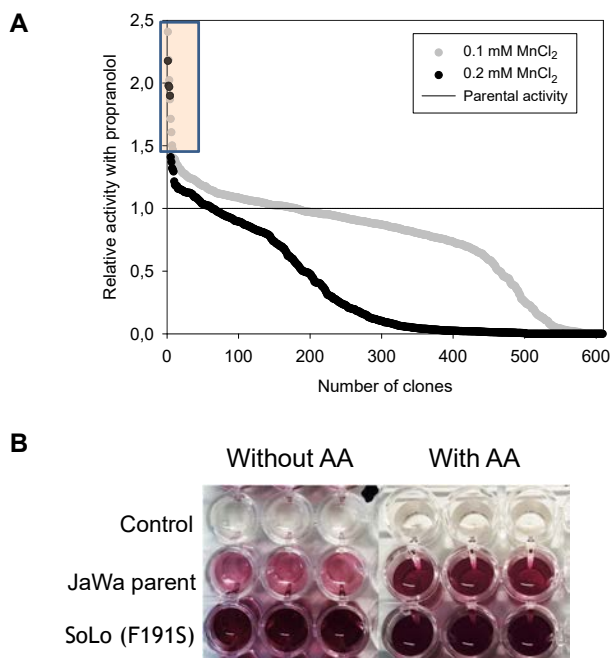**Figure S4: MORPHING libraries.** (**A**) Mutagenic landscapes obtained by MORPHING at different mutational frequencies. The activity of the clones is plotted in descending order and the solid line shows the activity of the parental type in the assay. The clones selected for rescreening are framed. (**B**) Activity of the parental UPO and the SoLo mutant in the 4-AAP assay, in the presence and absence of AA.
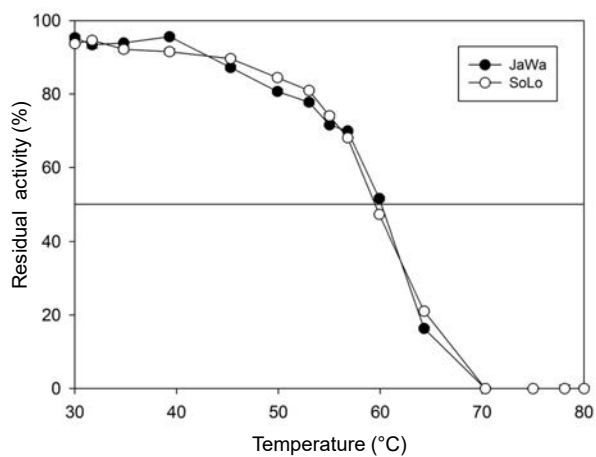
Figure S5: Kinetic thermostability of the JaWa and SoLo variants. The T_{50} was defined as the temperature at which the enzyme maintained 50% of its activity after a 10 min incubation. Each point and standard deviation is from three independent experiments.
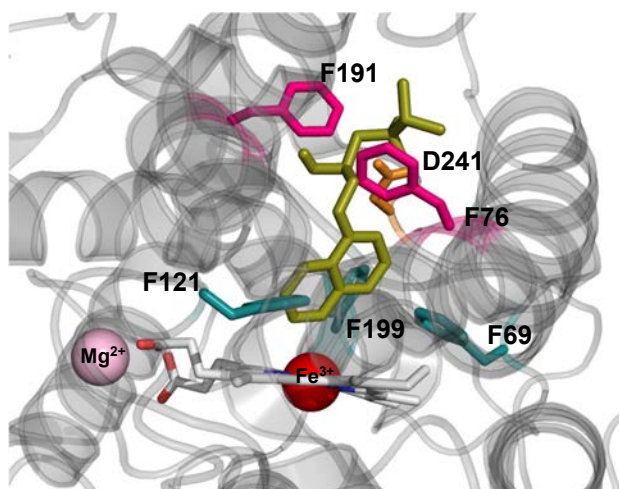
Figure S6: Positions subjected to saturation mutagenesis. The UPO structure is shown as a grey cartoon: the heme group is in CPK colors with Fe^{3+} as a red sphere, the Phe triad is in turquoise, the Phe191 and Phe76 pair delimiting the entrance to the heme is in pink, and Asp241 is orange. The structural Mg^{2+} is represented as a pink sphere and propranolol is in green.
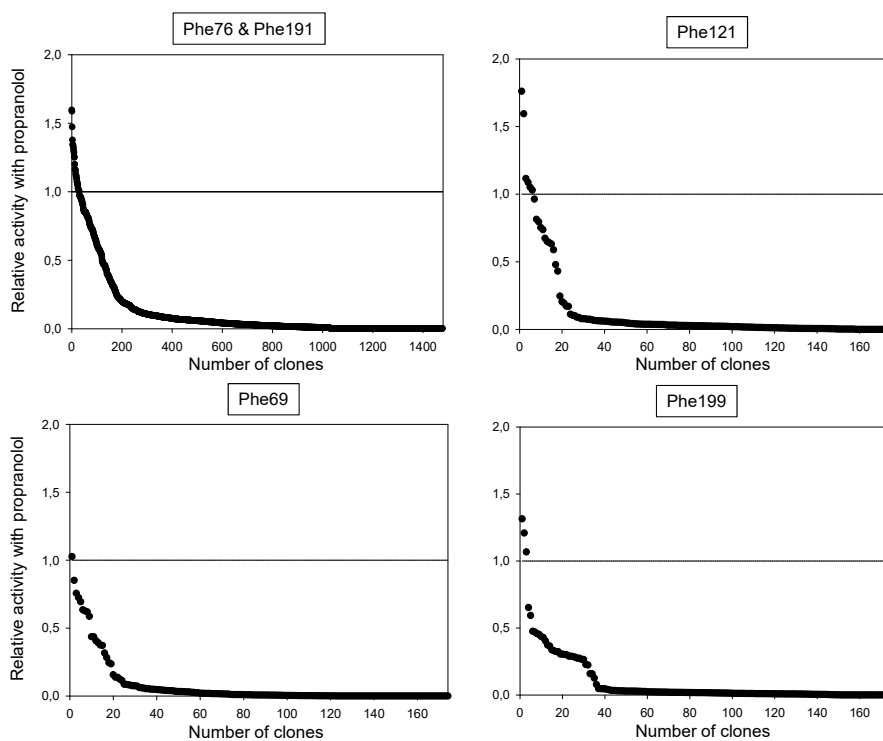
Figure S7: **Mutagenic landscapes of saturation mutagenesis libraries.** The activity of the clones is plotted in descending order; horizontal solid line indicates the activity of parental type.
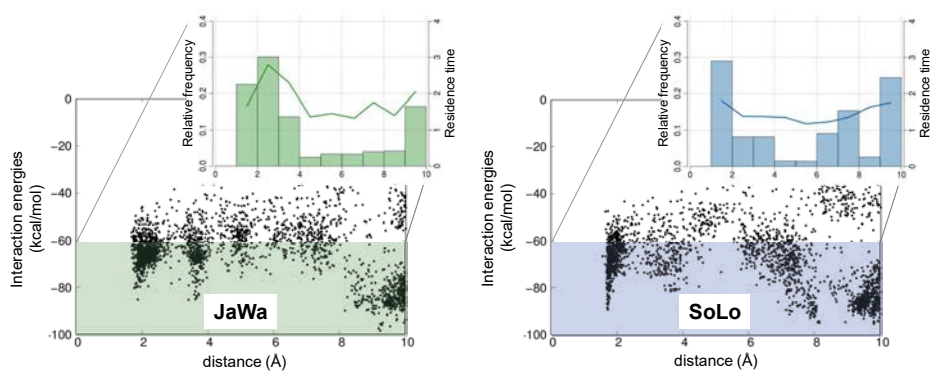
Figure S8: Computational analysis of 5´-OHP diffusion in JaWa and SoLo UPO variants. Interaction energies (in kcal·mol[-1]) *vs.* ligand distances (in Å) from PELE simulations with 5´-OHP in JaWa and SoLo UPO variants. The distances are between the reactive O atom in the heme compound I and the C_5 atom of 5´-OHP. The top-right inset bars show the relative distance distribution of those structures with energies below -60 kcal·mol[-1], along with a line indicating the average residence time of 5´-OHP in the binding site, calculated using binning widths of 1Å.

A

B
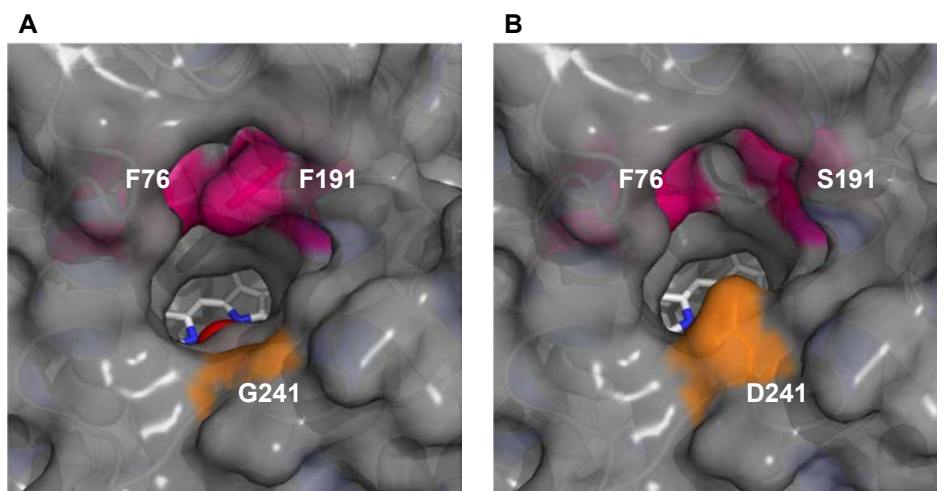
Figure S9: Differences in the heme access channel of the UPO variants. The Phe residues defining the entrance of the substrates are depicted in pink (**A**), with G241D in orange and F191S in pink (**B**). The G241D mutation is present in the JaWa and SoLo variants, whereas F191S appears only in SoLo (neither *Aae*UPO nor PaDa-I mutant contain such substitutions). Protein modeling based on the crystal structure of the evolved PaDa-I variant at a resolution of 1.5 Å (courtesy of Prof. Julia Sanz, IQFR-CSIC, Madrid, unpublished material).
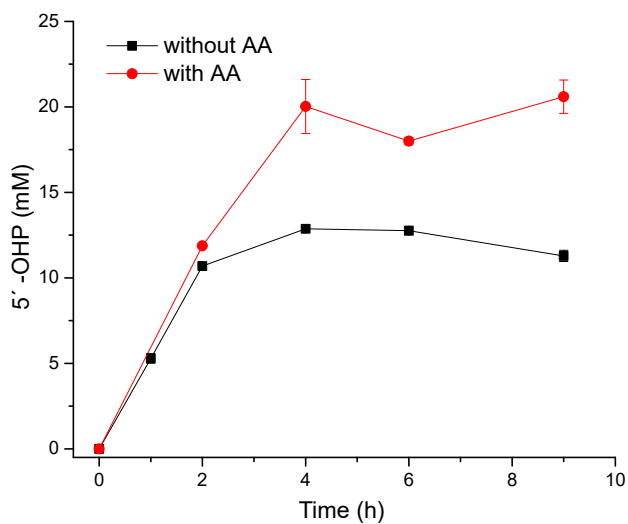
Figure S10: Semi-preparative production of 5′-OHP with SoLo mutant and *tert*-BuOOH. The reaction started by the addition of 5 mM *tert*-BuOOH and every hour, 5 mM *tert*-BuOOH was added to the reaction (further information can be found in the Materials and Methods section).
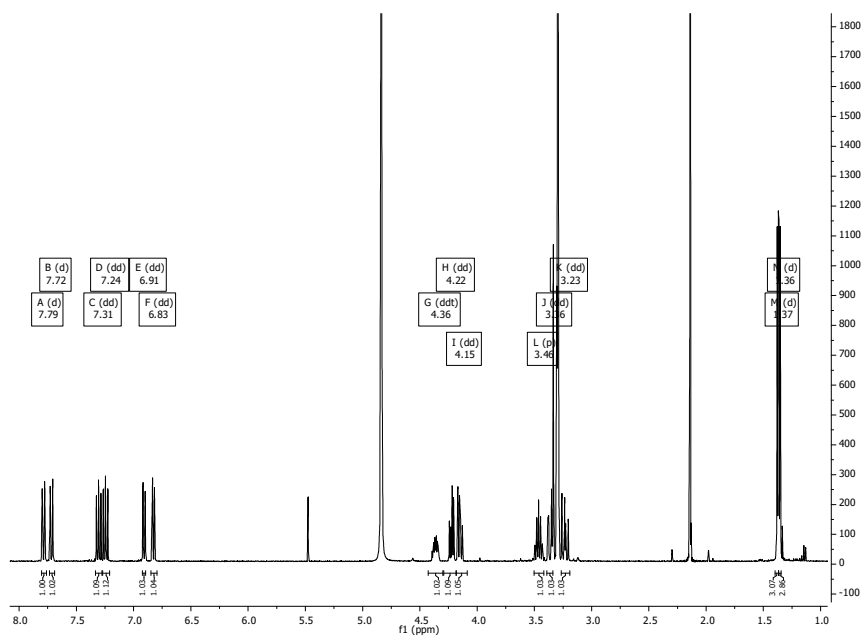
Figure S11: ¹H NMR of 5´-OHP after product isolation. 5´-OHP NMR (400 MHz, methanol-*d*4) δ 7.79 (d, 1H, *J* = 8.5 Hz), 7.72 (d, 1H, *J* = 8.5 Hz), 7.31 (dd, 1H, *J* = 8.5, 7.7 Hz), 7.24 (dd, 1H, *J* = 8.5, 7.7 Hz), 6.91 (dd, 1H, *J* = 7.7, 0.9 Hz), 6.83 (dd, 1H, *J* = 7.7, 0.9 Hz), 4.36 (ddt, 1H, *J* = 12.6, 5.4, 3.1 Hz), 4.22 (dd, 1H, *J* = 9.9, 5.0 Hz), 4.15 (dd, 1H, *J* = 9.9, 5.8 Hz), 3.46 (p, 1H, *J* = 6.6 Hz), 3.36 (dd, 1H, *J* = 12.7, 3.1 Hz), 3.23 (dd, 1H, *J* = 12.7, 9.6 Hz), 1.37 (d, 3H, *J* = 3.8 Hz), 1.36 (d, 3H, *J* = 3.8 Hz).

3.3 COMPUTATIONAL ENZYME REDESIGN

The idea of custom-made enzymes is as fascinating as challenging. It is fascinating in a way that if we had a better understanding of the molecular principles under enzyme catalysis, we should be able to tune up enzymes towards any preconceived task, by either modifying naturally existing proteins or using *de novo* scaffolds. Unfortunately, enzyme design is still far from trivial [153].

During the past two decades, DE has revolutionized the field of protein engineering, becoming the strategy of choice for tailoring properties of target proteins. However, DE comes with several problems related to the tiny sequence space that is able to sample and to some experimental biases [26, 147]. To overcome DE weaknesses, new computational strategies, such as enzyme rational or semi-rational redesign, are increasingly being used either as an alternative or complementary to DE. These techniques allow us to understand the molecular basis of enzymatic function and to use the obtained information to tailor enzymes toward desired properties, highlighting the importance of using knowledge-based methods for enzyme design.

In *Article VII*, computational methods have been used to obtain a complete biophysical description of a number of enzymatic reaction cascades that lead to the production of 2,5-furandicarboxylic acid (FDCA) in three different enzymes (*Aae*-UPO, HMFO, and AAO). Moreover, the use of the acquired information by molecular modelling simulations, along with the employment of a new semi-rational redesign *in silico* methodology, has enabled us to computationally engineer the *Aae*UPO enzyme towards an improved production of FDCA. Although the obtained variant stands only by theoretical basis and has not yet been experimentally tested, our results open a door of possibilities regarding the combination of *in silico* and *in vitro* methods for tailoring enzymes towards the production of high-value chemicals.

Publications presented in this section:

Article VII: <u>Cañellas M.,</u>Sancho F., Guallar V. and Lucas F. (2018) Towards the efficient production of FDCA: Molecular Modelling and *In Silico* Design of *Aae*-UPO, HMFO and AAO enzymes. (*Draft manuscript*)

3.3.1 *Article VII*: Towards the efficient production of FDCA: molecular modelling and *in silico* design of *Aae*UPO, HMFO and AAO enzymes. (*Draft manuscript*)

Marina Cañellas,[‡a] Ferran Sancho,[‡a] Victor Guallar[ab] and Fátima Lucas[c*]

[a] Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain

[b] ICREA, Passeig Lluis Companys 23, E-08010 Barcelona, Spain

[c*] ZYMVOL BIOMODELLING SL, C/ Almogavers, 165, 08018 Barcelona, Spain; E-mail: flucas@zymvol.com; Tel: +34617927468

[‡] These authors equally contributed to this work

Towards the efficient production of FDCA: Molecular Modelling and *In Silico* Design of *Aae*UPO, HMFO and AAO enzymes.

Marina Cañellas,[‡a] Ferran Sancho,[‡a] Victor Guallar[ab] and Fátima Lucas[c*]

[a] Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain

[b] ICREA, Passeig Lluis Companys 23, E-08010 Barcelona, Spain

[c*] ZYMVOL BIOMODELLING SL, C/ Almogavers, 165, 08018 Barcelona, Spain; E-mail: flucas@zymvol.com; Tel: +34617927468

[‡] These authors equally contributed to this work

_____

ABSTRACT

The growing interest in replacing fossil-based products with renewable chemical building blocks has focused researcher's attention towards the production of 2,5-furandicarboxylic acid (FDCA). Recently experimental work has proved possible the production of FDCA from 5-hydroxymethylfurfural (HMF) by a reaction cascade that combines the action of aryl-alcohol oxidase (AAO) and an unspecific heme peroxygenase from *Agrocybe aegerita* (*Aae*UPO). Additionally, hydroxymethylfurfural oxidase (HMFO) has been reported to catalyze the whole reaction. In this work, computational methods have been used to characterize at the molecular level of detail the before-mentioned FDCA-producer enzymatic systems. *In silico* results have enabled us to rationalize the AAO inability to produce FDCA by itself, the slow reactivity of *Aae*UPO, and the reasons behind the improved activity of an HMFO double mutant. Moreover, valuable molecular-based information extracted from simulations has enabled us to computationally engineer *Aae*UPO enzyme towards the improvement of FDCA production.

INTRODUCTION

During the latest years, the awaited decrease in petroleum reserves together with the need to reduce greenhouse gas (GHG) emissions, have created the necessity to replace classical fossil-based sources. In this way, a growing interest has appeared in the use of renewable carbon sources for the production of biological-based chemicals, being lignocellulosic biomass the perfect candidate due to its abundance, economic feasibility and sustainable nature.[1-3] The main goal of lignocellulose biorefineries is to produce biofuels, such as ethanol and biodiesel, to replace a part of the large amount of gasoline and diesel used on transport worldwide (~14 millions of barrels/day in 2016).[4] However, despite this enormous demand, fuel is still cost effective compared to biofuels. For this reason, the development of processes to produce valuable biochemicals is also a priority for biorefineries in order to enhance revenues and repurpose existing underemployed infrastructures.[5-7]

In 2004, the United States Department of Energy (DOE) released a report outlining 15 targets that could be of interest for the manufacture of bioproducts.[8] Based on this report, some of these "top chemicals" have been the subject of intense research over the last decade, leading to a significant progress in this area.[5] One compound identified to be of remarkable importance is 2,5-furandicarboxylic acid (FDCA).[1,5,6,8] Due to its structural analogy with terephthalic acid, a precursor of polyethylene terephthalate (PET), FDCA has been proclaimed as a green alternative to this petroleum-based chemical. FDCA is also used in the production of polyethylene furanoate (PEF) polymer,[9,10] which has the advantage of being renewable and biodegradable while showing similar properties to other polymers derived from terephthalic acid.

Several chemical methods to obtain FDCA from sugars have been studied. However, most procedures require high temperature, high pressure, metal salts and organic solvents which render the process polluting and expensive.[13] For this reason, enzymatic conversion seems to be a suitable alternative as it can be performed at lower temperatures and pressures (lowering capital expenditures), while offering superior chemo-, regio- and enantioselectivity, and with the plus of being environmentally friendly.[14] FDCA can be synthesized from sugars, generally fructose. First, fructose is dehydrated and converted into HMF, a process that requires acid catalysis and relatively high temperatures.[12] In the second step, HMF is oxidized into FDCA, a reaction that can proceed by two different 3-stepped routes: via 2,5-diformylfuran (DFF) or via 2,5-hydroxymethylfurancarboxylic acid (HMFCA). Both metabolites are precursors of 2,5-formylfurancarboxylic acid (FFCA), which is then oxidized to FDCA. The whole reaction scheme is shown in **Fig. 1**.

Nevertheless, the biochemical production of FDCA from HMF is not well established and biocatalytic approaches are scarce.[15,16,41] Bioproduction of FDCA by a single enzyme is complicated since most catalysts are restricted to either alcohol or aldehyde oxidation and product generation involves acting on both groups. Heretofore, only one enzyme has been described to oxidize HMF giving FDCA as unique product, named HMF oxidase (HMFO). However, HMFO enzyme shows poor activity towards FFCA intermediate, limiting the overall process' efficiency.[17,18] To overcome this problem, Dijkmen *et al.* proposed two mutations (V367F and W466F) on the HMFO enzyme binding site, incrementing by 1000-fold the enzymatic catalytic efficiency of HMFO wild-type.[19]

Recently, the full enzymatic conversion of HMF into FDCA was reported using a cascade reaction with two different fungal enzymes: aryl-alcohol oxidase (AAO, EC 1.1.3.7), which shares the active site's structural architecture of HMFO enzyme and is secreted by white-rot basidiomycetes, and an unspecific peroxygenase (UPO, EC 1.11.2.1) from *Agrocybe aegerita*.[20] AAO is able to oxidize HMF to FFCA (via DFF) in a few hours. However, it is unable to directly oxidize FFCA and thus is not retrieving the final product, FDCA. On the other hand, the UPO enzyme is capable of catalyzing the whole reaction via HMFCA, including the FFCA oxidation, despite the fact that each step is progressively slower. Therefore, the protocol proposed by Carro et al. involves the conversion of HMF to FFCA by AAO with the concomitant production of H_2O_2, which is then used by the UPO to produce FDCA from FFCA. In this protocol, H_2O is the only by-product formed and O_2 the only co-substrate required (**Fig. 1**). Nevertheless, the overall process is slow due to a very inefficient conversion of FFCA to FDCA by UPO.

The goal of the present work is to understand the molecular details that govern the enzymatic transformation of HMF into FDCA in both AAO and UPO enzymes, and in particular, dissect the reasons for the reduced activity on FFCA. To this aim, each oxidation step of the two different reaction routes was studied using computational techniques. DFF route (blue colored in **Fig. 1**) was investigated with AAO and HMFCA route (green colored in **Fig. 1**) with UPO. Additionally, DFF route was also studied with HMFO enzyme and its Val367Arg/Trp466Phe double mutant (HMFO-*mut*) (**Fig. 1**). Energy profiles and substrate-ligand binding modes of each system were determined by Protein Energy Landscape Exploration (PELE) software[21], a structure-based computational technique. Results show that computational simulations can provide a rational explanation, at a molecular level, of the differences in yield and regioselectivity observed experimentally. Additionally, all this valuable molecular-based information has enable us to perform the *in silico* engineering of AAO and UPO enzymes towards the theoretical improvement of

FDCA production, by using both rational design and an *in silico* evolution proto-col.[36]
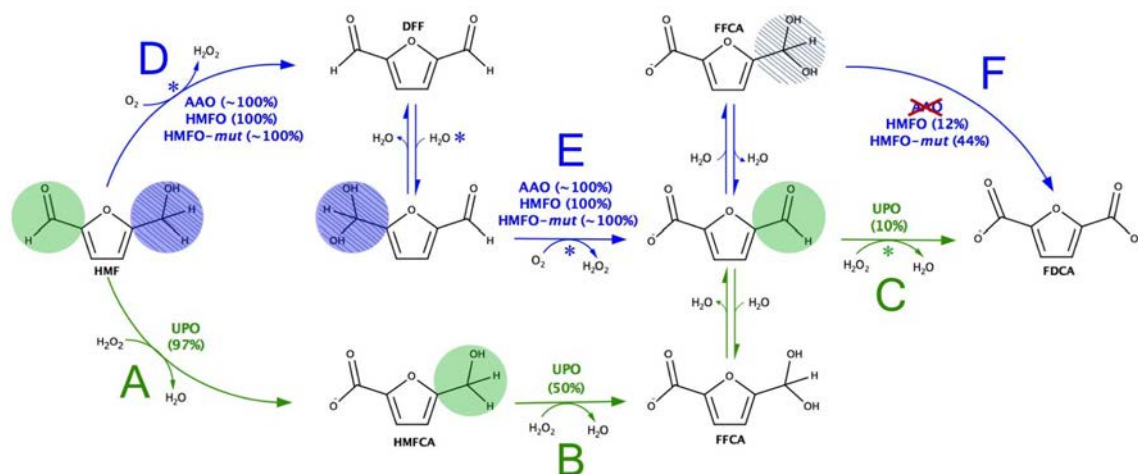
Figure 1. Schematic representation of HMF conversion into FDCA by two different routes: via DFF (catalyzed by AAO, HMFO and HMFO-mut enzymes, colored in blue) and via HMFCA (catalyzed by UPO enzyme, colored in green). Functional groups that are oxidized by UPO in each step of the reaction are circled in green, while functional groups oxidized by AAO are circled in blue and blue-stripped for HMFO/HMFO-mut. Reaction steps marked with an asterisk are those involved in the protocol proposed by Carro *et al.* to obtain FDCA.[20] For each step, the conversion rate is shown considering HMF as the reactant in a 24h time period. To simplify the reaction identification in the overall manuscript, each step has been given a letter (from reaction A to F).

METHODS

System Setup

The starting structure for computational simulations of *Agrocybe Aegerita* UPO was the 2.19 Å resolution crystal (PDB code 2YOR) obtained by Piontek and coworkers.[22] Crystal structures with PDB code 3FIM and 4UDP were used respec-tively for simulations with AAO and HMFO. Simulations were performed at pH 7, reproducing the experimental conditions, which has been found to be a good com-promise for the three proteins in terms of performance and stability.[20] Protein struc-tures were prepared accordingly using Schrodinger's Protein Preparation Wizard PROPKA module[23,30,31] and H++ web server (http://biophysics.cs.vt.edu/).[24] In

UPO enzyme, all histidines were δ-protonated with the exception of His82 (ε-protonated), His118 and His251 (double-protonated), and all acidic residues were deprotonated, except Asp85 that was modeled in its protonated state. For AAO enzyme, histidines 91, 313, 387 and 398 were ε-protonated, 190, 502 and 546 δ-protonated and the remaining histidines were doubly protonated. All acidic residues were deprotonated except Glu389, according to previous studies.[29] Finally, for HMFO enzyme, histidines 51, 287 and 307 were ε-protonated, His149 was fully protonated and the remaining histidines were δ-protonated.

All the ligands were optimized with Jaguar[25] with the density functional M06 with 6-31G** basis set and a Poisson Boltzmann Finite (PBF) implicit solvent in order to obtain their electrostatic potential derived atomic charges. These include HMF, DFF, FFCA (deprotonated, pKa=3.09) and HMFCA (deprotonated, pKa= 3.11). Note that for AAO and HMFO simulations, FFCA and DFF were modeled on their hydrated form. All pKa values were collected from www.chemical-ize.org.[32] On the other hand, the heme cofactor in UPO was modeled as compound I after being fully optimized in the protein environment with quantum mechanics/molecular mechanics (QM/MM) using QSite.[26] The same method has been employed to optimize the flavin adenine dinucleotide (FAD) cofactor in both AAO and HMFO enzymes. The resulting atomic charges were used for classical simulations.

Protein Energy Landscape Exploration (PELE)

To investigate the differences observed experimentally on yield and regioselectivity between the distinct reactions, we have used PELE software to perform substrate migration simulations on all enzymes with their corresponding substrates. PELE is a Monte Carlo based algorithm that produces new configurations through sequential ligand and protein perturbations, side chain prediction and minimization steps. New configurations are then filtered with a Metropolis acceptance test, where the energy is described with the all-atom OPLS2005 force field[27] and a surface generalized Born implicit solvent.[28] In this way, it is possible to locate and characterize local and global minima structures for the most favorable protein-ligand interactions. PELE has been used successfully in many ligand migration studies with both small and large substrates.[33-35]

Once initial protein structures were prepared, the optimized ligands were placed manually in identical positions at the entrance of each corresponding protein binding pocket, and PELE simulations were performed. In this work, PELE was set up first to drive the ligands inside the protein by reducing the distance between the

ligand's center of mass (COM) to the enzyme binding site (heme reactive oxygen in UPO and FAD's reactive nitrogen in AAO/HMFO/HMFO-*mut*) up to about 5 Å. Then, once the ligand was inside the active site, it is allowed to explore it freely (with no restraints). The results presented are based on, at least, 160 48-h trajectories for each system.

In silico evolution protocol

The computational protocol applied in this work to perform *in silico* enzyme evolution in UPO has been previously proposed by Monza *et al.*[36] and successfully applied for the design of POXA1b laccase by Giacobelli *et al.*[37] This protocol demands a scrupulous selection of the initial structure, which must embody a catalytic substrate positioning, and that will determine the outcome of the computational design. In this way, the initial structure for the *in silico* design was selected from the previously performed FFCA substrate diffusion PELE simulations in the UPO enzyme, picking those structures with optimal catalytic distances. Once the initial structure was selected, every residue within 5 Å from the substrate was selected as potentially mutable. HotSpot Wizard 2.0 webserver[38] was used to discard those residue positions that are directly involved in protein catalysis or located at evolutionary-conserved regions. Finally, only 10 residue positions, including 69, 73, 77, 76, 123, 188, 192, 193, 195 and 199, were mutated to all the possible remaining 19 amino acids. After mutation, the ligand was perturbed with 0.1 Å translations and 0.1 rad rotations whereas the protein backbone was subjected to 0.75 Å ANM displacements, and all the residues within 5 Å from the substrate were energy minimized. For each mutant, a 2-h simulation was performed, obtaining as a result the enzyme-substrate interaction energy that was used to rank the 190 mutants. Finally, FFCA diffusion was performed as explained in the Protein Energy Landscape Exploration (PELE) section for the top-5 mutants. The results presented here are based on 160 48-h trajectories.

Molecular Dynamics and MDpocket.

To study the structural changes caused by mutations, 100 ns MD simulations were performed with UPO-*mut* structure using Desmond[39]. To prepare the system for molecular dynamics, previously prepared UPO-*mut* system was placed inside an orthorhombic box containing SPC explicit waters and ions to neutralize the system at a concentration of 0.15 M NaCl. From the solvated system, MD simulation was run with the following parameters: the OPLS-2005 force field, the temperature was regulated with a Nose-Hoover chain thermostat with a relaxation time of 1.0 ps, the pressure with the Martyna-Tobias-Klein barostat with isotropic coupling

and a relaxation time of 2.0 ps, and finally, the production phase was run over the course of 100 ns using the NPT canonical ensemble at 300K. From MD simulation, structures at every 0.1 ns were extracted and used for volume pocket calculation with MDpocket[40], a fast and open-source tool for protein pocket (cavity) detection on molecular dynamic trajectories or other conformational ensembles.

RESULTS AND DISCUSSION

To investigate the differences observed in the enzymatic catalysis of each system, substrate migration simulations were performed for each reaction step in both paths: via DFF (blue, D-F reactions in **Fig. 1**) and HMFCA (green, A-C reactions in **Fig. 1**) using PELE. From the simulations, energy profiles and binding modes were obtained. All energy plots are depicted as the variation of the interaction energy versus the catalytic distance between the ligand and the enzyme' binding site. In this way it is possible to identify the main energy minima along the substrate's migration from the protein's surface to the binding site. Moreover, rational design and *in silico* enzyme evolution protocols were synergistically used to model the UPO protein towards the more efficiently production of FDCA from FFCA in an attempt to theoretically enhance the whole reaction.

Unspecific peroxygenase (UPO) VIA HMFCA – reactions A, B, and C.

Reaction A. According to the results obtained by Carro *et al.*,[20] HMF is oxidized into HMFCA by UPO protein with almost 100% conversion in 24h. From the molecular simulations, where we have explored the entrance of all substrates from the protein surface to the heme active site (**Fig. 2**), we can see how HMF can be placed in two important minima in the UPO binding site (**Fig. 2A**). In the first one (cyan colored in **Fig. 2A**), the ligand is in an optimal position for reaction, with the aldehyde's hydrogen at ~2.5 Å from the heme reactive oxygen (O-Heme). This position is stabilized by two hydrogen bonds with Thr192 and Phe121, as shown in the molecular representation of **Fig. 2A**. However, there is a competing conformation (colored in green in **Fig. 2A, Figure S1**), with the hydroxyl group oriented towards the heme (thus the distance of the reacting oxygen is over 6Å away from the iron). In this second minimum, the HMF hydroxyl group forms a hydrogen bond with the O-Heme that is conserved along the simulation. Due to this interaction, hydrogens H_2 and H_3, which are adjacent to the hydroxyl group, cannot approach the O-Heme to a reactive position. To confirm this hypothesis, we have computed the average distances between the different hydrogens (H_2, H_3 and H_6) and the O-Heme in structures corresponding to this minimum. Results show that the average distance between H_6 and O-Heme is 1.77 Å, compared to 4.27 Å for H_2 and 3.55 Å for H_3,

providing evidence that support our hypothesis. Therefore, we propose that, be-cause of this conserved interaction, UPO enzyme is not able to oxidize HMF to DFF and can only proceed via HMFCA formation. This reaction is nevertheless slow (24h), possibly due to the presence of an unproductive competing minimum.

Reaction B. In the case of the HMFCA reaction, a different energy landscape is found where strong energy minima are located at ~12 Å from the heme (**Fig. 2B, orange colored**). The existence of these minima (not seen for HMF) seems to be due to the negatively charged nature of the ligand's carboxylate group. Visual in-spection of these structures reveals a main position where the ligand is placed in the entrance of the heme channel (**Fig. S2**). Interestingly, this particular ligand po-sitioning leads to a reduction of the diameter of the heme channel entrance (sup-plementary computational simulations have been performed with FFCA substrate to show the same event, see **Figure S5**). This narrowed entrance, along with the consequent difficulty of the ligands to reach the active site, may explain not only the slower catalysis shown by UPO for HMFCA compared to HMF, but also the overall UPO enzyme slow catalysis, since HMFCA product formation may also interfere with the protein, reducing its overall activity by noncompetitive inhibi-tion. This is in a good agreement with experimental results, which show a de-creased oxidation rate of HMF over time (72% in 8h versus 97% in 24h).[20]

However, despite the narrowed heme path entrance, the substrate is still able to enter the heme site and adopt favorable orientations for the reaction. As it can be seen in **Fig. 2B**, there are two different minima around -35 kcal/mol, both of them with the hydroxyl group (reactive functional group) close to the binding site. In one minimum (**Fig. 2A, green colored**), this group is oriented toward the O-Heme atom forming a hydrogen bond with it, and impeding H_2 and H_3 to approach the O-heme, and thus, disabling the ligand to react (**Figure S3**). On the contrary, the structures corresponding to the other minimum (**Fig. 2B, cyan colored**), the hydroxyl is pointing away of O-Heme, and hence, the substrate's H_2 and H_3 atoms are able to approach O-Heme atom within an appropriate distance to react. See **Fig. 2B** for a molecular representation of this minimum.
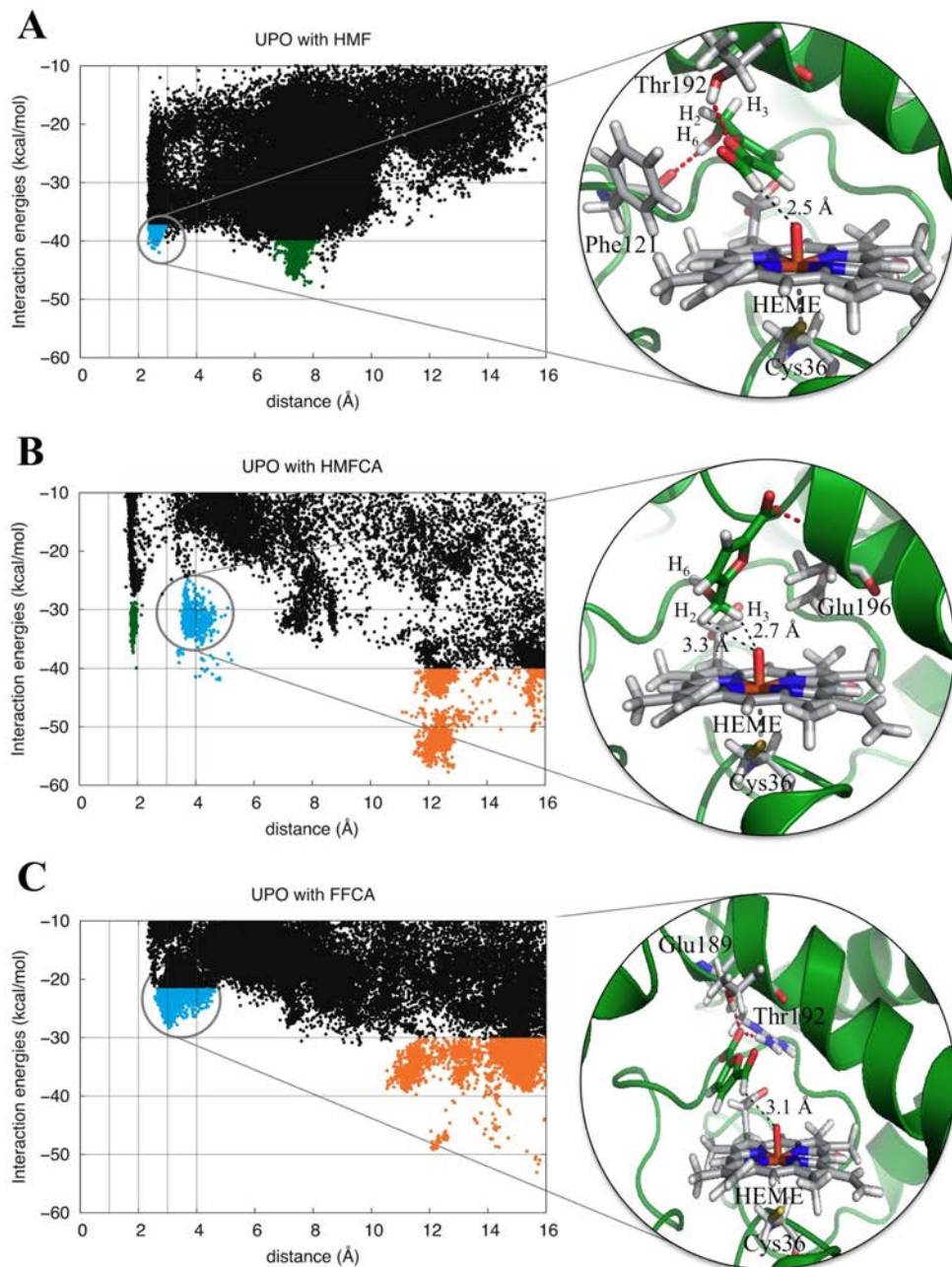
Figure 2. The plots on the left are the representation of the interaction energy of each PELE structure against its reactive distance in UPO from *A. aegerita* for HMF (A), HMFCA (B) and FFCA (C). This distance (Å) is measured between the reactive O atom in the H_2O_2-activated heme (compound I) and reactive aldehyde hydrogen (A,C) or the hydrogen of the hydroxyl group (B). The reactive minima are colored in cyan, while non-reactive are colored in green (in the active site) or orange (at the entrance). On the right, a representative

structure for the specified minimum is shown. Hydrogen bonds between the ligand and nearby protein residues are marked in red, and the distance between the heme oxygen atom and its closest ligand hydrogen atom are marked in black. The heme molecule, the heme axial residue (Cys36) and the protein residues that interact with the substrate are colored in grey, and the substrate itself is colored in green.

Reaction C. FFCA oxidation is of great interest since it is the rate limiting process for the production of FDCA via HMFCA. In **Fig. 2C** it can be seen that there are several strong minima in the entrance of the heme cavity (**Fig. 2C, orange colored**) as observed in HMFCA results (**Fig. 2B, orange colored**). Notice that FFCA is also negatively charged. An example of the ligand's binding mode on these minima can be seen in **Figure S4**. Similarly to HMFCA, the minima at the entrance of the heme channel induce a narrowing of this channel, hindering the access to other ligands. To confirm this event, additional PELE calculations have been performed with the FFCA ligand. For these simulations, we placed two FFCA molecules in the two main minima observed on the heme entrance (orange colored of Fig. 2C) and we kept their position restrained along the simulation of a new FFCA substrate migration with PELE from the surface of the protein to the binding site. By doing this, we could observe the narrowing of the heme path entrance by ~5 Å (**Figure S5**) and the consequent difficulty of the ligand to reach the binding site. Once the ligand is in the heme site, there is a less favorable minimum with the reactive aldehyde group near the O-Heme atom, which indicates that the reaction can still be carried out (**Fig. 2C, cyan colored**). However, the distance between the O-heme and the hydrogen of the aldehyde group is relatively large (>3 Å, see **Fig. 2C** for an example of binding mode), which hinders the reaction.

UPO *in silico* design: an improved variant for FDCA production.

As explained previously, we hypothesize that there are two main reasons why UPO enzyme has an inefficient FDCA production: the presence of global energy minima at the entrance of the heme path and a non-optimal substrate positioning in the binding site. To overcome the first problem, we visually inspected the structures from those minima and realized that in most of the cases, FFCA substrate was hydrogen bonded to Thr242 or Ala317 residues. In an attempt to remove such minima and to ease the entrance of the ligand to the binding site, we propose two mutations: Thr242Asp and Ala317Pro, which we have observed computationally to reduce these minima (green energy profile in **Fig. 3**). Moreover, the *in silico* evolution protocol explained in the Methods Section was applied to the UPO enzyme in order to improve FFCA binding, resulting in three highlighted mutations: Phe121Lys, Ala77Ser, Ala77Gln and Ala193Ser. Different mutation combinations were studied using PELE for the diffusion of FFCA substrate. The best design was

shown to be the combination of the two mutations at the entrance of the heme path, Thr242Asp and Ala317Pro, and Ala77Ser and Ala193Ser mutations in the binding pocket. As shown in **Fig. 3**, the new variant (UPO-*mut1*) has an improved FFCA binding site positioning when compared to the parental UPO, having a pronounced minimum in the binding site (around -40 kcal/mol) where the ligand is in an optimal catalytic position (<2.4 Å), along with a smoother entrance to the binding site.
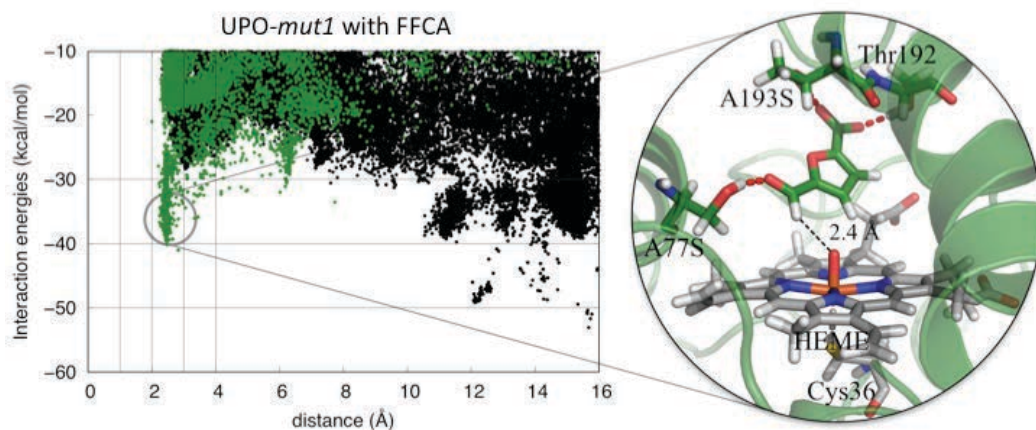

Figure 3. UPO-*mut1* variant with FFCA substrate. On the left there is a representation of the interaction energy of each PELE structure against its reactive distance for the *wt* UPO (colored in black) and for the UPO-*mut1* variant (colored in green). Reactive distance (Å) is measured between the H_2O_2-activated heme (compound I) and the substrate reactive aldehyde hydrogen. On the right, a representative structure for the specified minimum is shown. Hydrogen bonds between the ligand and nearby protein residues are marked in red, and the distance between the heme oxygen atom and its closest ligand hydrogen atom is marked in black. The heme molecule and the heme axial residue (Cys36) are colored in grey, and the substrate itself and the protein residues that interact with the substrate are colored in green.

To further characterize the proposed UPO variant and to obtain structural insights into the changes caused by mutation, 100 ns molecular dynamics (MD) were performed with UPO-*mut1* structure. Results show that the overall UPO-*mut1* structure does not suffer from big conformational changes over 100 ns MD simulation (heavy atom RMSD ~2Å). However, comparison between 1 ns and 100 ns MD structures show a possible widening of the propionate channel (**Fig. 4A**) caused by mutation. To further study this hypothesis, MDpocket method was used to track heme's and propionate's cavity volume changes along the MD trajectories. As shown in **Fig. 4B**, results do not show a significant widening of the propionate cavity along simulation, discarding previously mentioned hypothesis.
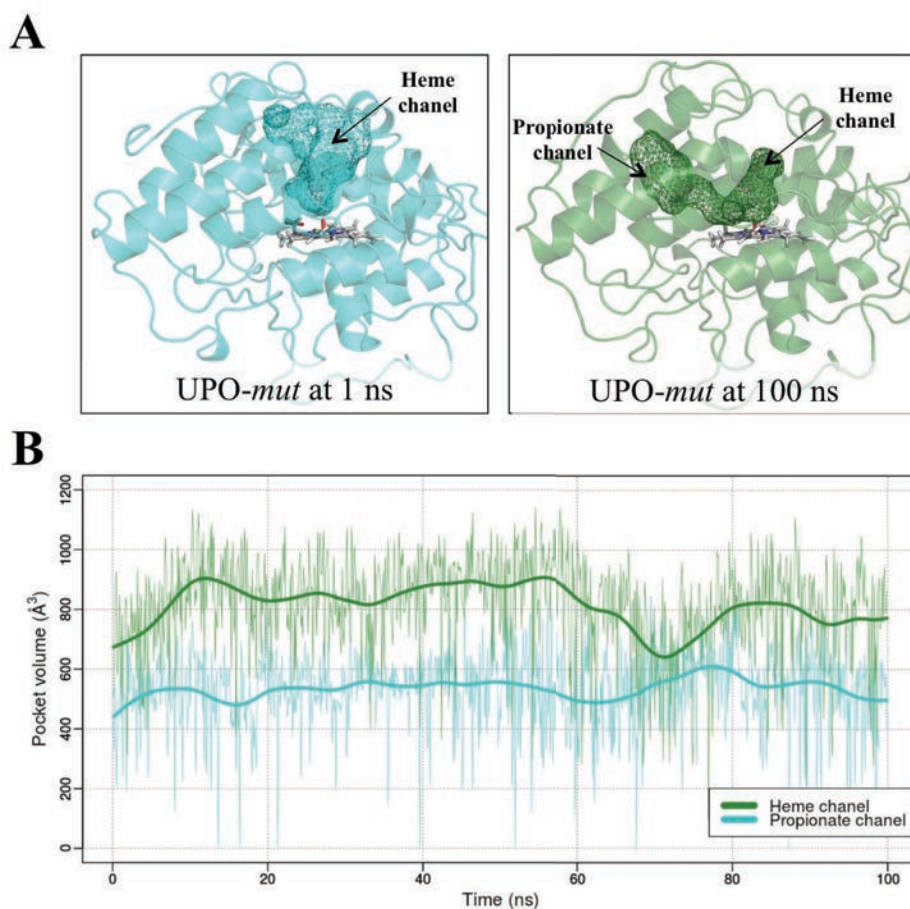
Figure 4. Mutation-induced structural changes along 100 ns MD. A – UPO-*mut* structure at the beginning (0 ns, cyan colored) and at the end (100 ns, green colored) of the MD simulation. Heme and propionate pocket volume are represented as a mesh surface. **B –** Heme (green colored) and propionate (cyan colored) pocket volume tracking along 100 ns MD simulations for UPO-*mut*, calculated with MDpocket every 0.1 ns. Smoothed volume of the pocket is shown in thick lines.

Aryl alcohol oxidase (AAO) VIA DFF – reactions D, E, and F.

As explained in the introduction, is not possible to produce FDCA via DFF through AAO catalysis. Despite the fact that HMF conversion to DFF, and DDF to FFCA occur very rapidly (~100% conversion in less than 4h), the final step (conversion of FFCA to FFDA) is not possible. Nevertheless, the non-necessity of H_2O_2 as a co-substrate and the fast conversion of the two first steps generated a huge interest in understanding the reasons why the FDCA production catalyzed by

AAO is not possible, which opens the possibility for tailoring the protein to overcome this limitation.
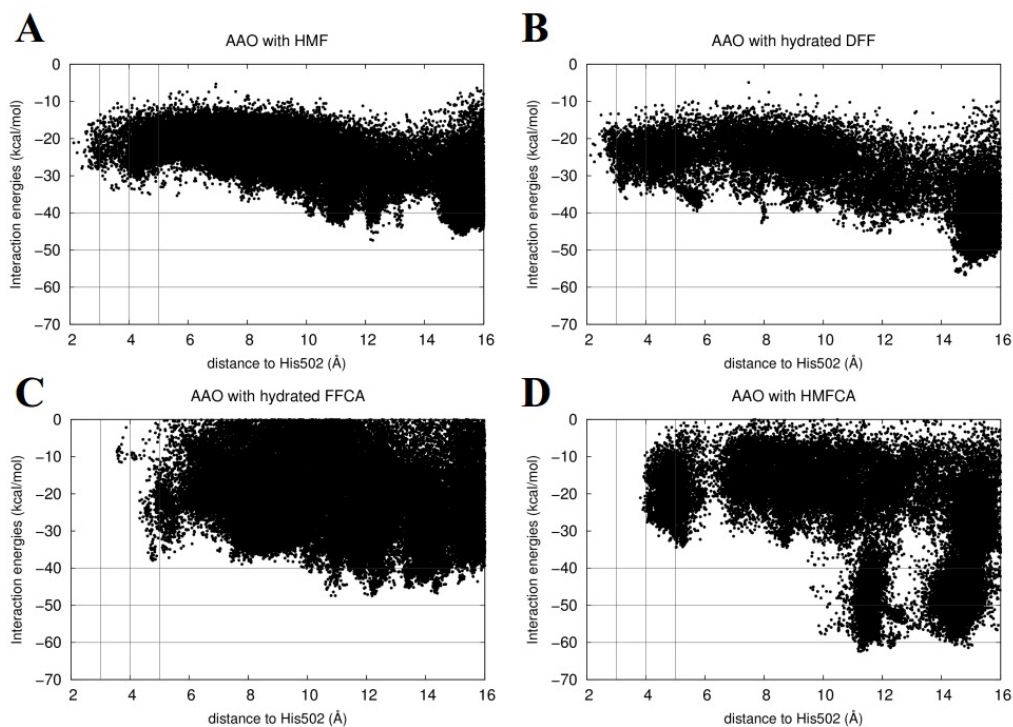
Figure 5. Interaction energies vs. ligand distances from PELE simulations in AAO from *Pleurotus eryngii.* The distances shown (Å) are measured between the reactive N atom in the His502 and the alcohol hydrogen atom from the ligand. The plots correspond to the migration of different ligands: **A** - HMF, **B** - DFF, **C** - FFCA and **D** - HMFCA.

In order to inspect the differences in reactivity between the substrates, particularly the incapability of FFCA to be oxidized by this protein, and following the protocol explained in the methods section, ligand migrations of HMF, DFF, FFCA were performed in AAO. Based on previous evidence,[29] the reaction mechanism is expected to follow a concerted nonsynchronous proton transfer followed by hydride transfer and thus, the distance between the hydrogen from the ligand's alcohol and His502 nitrogen is the reaction coordinate to consider. Therefore, the interaction energy along the simulations has been represented in **Fig. 5** as a function of this distance. A threshold of 3Å was established in order to consider that the ligand is close enough for the proton extraction to take place. As shown in **Fig. 5**, both HMF and DFF reach distance values lower than 3Å, which are appropriate for the

reaction, while FFCA cannot get closer than 4Å. Based on this difference we hypothesize that FFCA is not converted by AAO due to the fact that it cannot approach the active site to an appropriate catalytic position.

Additional simulations with the HMFCA substrate, an intermediate product obtained in the UPO path that has been proven experimentally to be non-reactive in AAO, were also performed. Interestingly, its energy profile shows similar results to FFCA, stalling at 4Å of the previously mentioned distance. Inspecting the structure of both ligands, it can be noticed that they share a common characteristic, which is a negatively charged carboxylic group at the working pH. By analyzing the energy profiles, the structures corresponding to the minimum located at 4Å for both ligands have been extracted, and a representative conformation for FFCA ligand is shown in **Fig. 6**. As evidenced, the negatively charged group interacts with many residues along the path, such as Tyr92, His398, Gln395 and Arg403, making the substrate unable to reach the active site. These results provide further indications that the negative charge of the carboxylate could be the responsible for the null activity of both ligands in AAO.
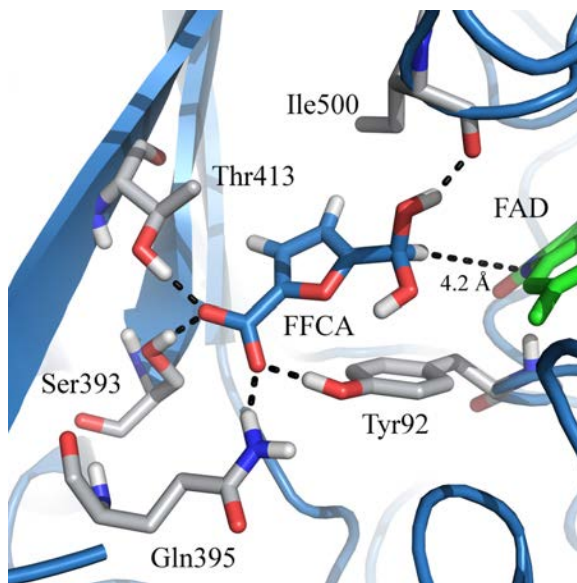
Figure 6. Structural representation of the FFCA ligand conformation when the catalytic distance is around 4 Å. Strong interactions of the carboxylic group of the ligand with polar residues can be appreciated.

Hydroxymethylfurfural oxidase (HMFO) VIA DFF – reactions D, E, and F.

The study of HMFO mechanism is of high interest, since it was the first enzyme reported to be able to catalyze, on its own, the entire reaction chain from HMF to FDCA via DFF, and also, because it shares a similar structural architecture of the active site with AAO. The same set of ligands (HMF, DFF, FFCA) have been simulated in the HMFO system with PELE for both the native and the Val367Arg/Trp466Phe variant (HMFO-*mut*). As previously stated, HMFO-*mut* is able to catalyze the formation of FDCA with around 1000-fold increment compared to the wild type.
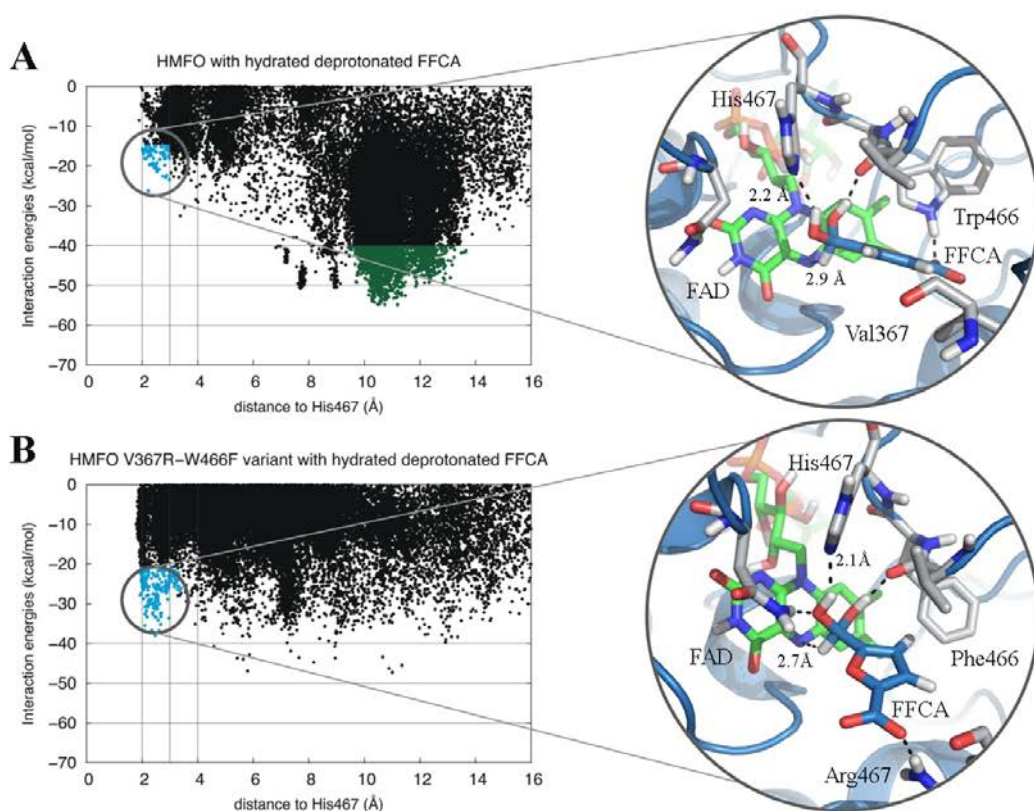
Figure 7. Two dimensional interaction maps showing the interaction energy vs. the distance between FFCA and the His467 for the native HMFO (A) and HMFO-mut (B). A representative structure from the catalytic minimum was taken from the native protein where it can be seen the interaction between Trp466 and the ligand (A). For HMFO-mut, a representative structure of the active site was extracted to show the contribution of the Arg367 in the positioning of the ligand.

The energy profiles for the FFCA migration in both HMFO and HMFO-*mut* are represented in **Fig. 7**. It is shown how for the native protein (**Fig. 7A**), the ligand is able to contact the active site with reasonable catalytic distances for the oxidation to occur (< 3 Å), but the catalytic position is not energetically favored. Around 10 Å from the reacting hydrogen of the ligand to His467, a strong minimum is identified and its interaction energy difference with the reactive position is considerably large. In practice, this suggests that the active conformation is not very populated which translates in a poor reactivity towards the oxidation of FFCA. On the contrary, HMFO-*mut* is able to catalyze the last reaction step more efficiently. The same energy profile is plotted for the FFCA migration in the double mutant variant (**Fig. 7B**). In this case, it can be seen how the substrate is able to approach the active site with good catalytic distances and energetically favored positions when compared to the rest of the trajectory.

According to the results represented in **Fig. S6**, the two mutations introduced in HMFO induce minimal changes to the interaction of the protein with HMF and DFF, considering that the energy profile is barely altered. Both ligands can still reach the active site close enough to permit the ligand's reaction. Yet, a significant improvement is actually obtained in the limiting step, which corresponds to the FFCA oxidation as stated previously. This enhancement is due to the alterations introduced by the mutations: in the first place, with the substitution of Trp466 for a phenylalanine, a smaller residue fills its place and the polar -NH group is no longer present. As presented in **Fig. 7**, this mutation causes adjustments in the interaction between FFCA and the protein, turning into a loss of several hydrogen bonds with respect to the native protein. Secondly, the inclusion of an arginine in substitution of Val367 (located in front of the FAD cofactor) adds a positive charge that has a strong interaction with the negative charge of the ligand, orienting the alcohol group towards the FAD and thus, helping the appropriate placement of the ligand for the oxidation. As a result, the catalytic distances are shortened by 0.1-0.2 Å, which can be expected to reduce the energy barrier of the proton and hydride transfers.

CONCLUSIONS

There is an imperious interest on finding bio-based alternatives for the production of FDCA from HMF, and in particular, the idea of a single system able to catalyze this triple oxidation reaction leading to HMF valued chemical seems extremely appealing. However, up to now, only few systems have been proven to perform the whole transformation, and in most of the cases, reaction from FFCA to FDCA has still a really poor conversion. Here, the study of these enzymatic transformations using molecular modelling techniques has been able to dissect, at the molecular level, the reasons underneath the activity at each step of the reaction, and to visualize which would be the perfect scene for this process to occur. In this way, PELE simulations have shown that for both UPO and AAO enzymes there are two main reasons for the processes to have scarce or null conversions: the inability of the ligand to reach the binding site with catalytic distances and the existence of an energy minima at the entrance of the binding site channel that hinders the ligand arrival to the binding site. Moreover, thanks to the valuable information obtained from the molecular modelling simulations and to the employment of a new methodology for *in silico* enzyme evolution, we have been able to computationally engineer the UPO enzyme. And, despite the fact that this variant relies only on a theoretical basis, it opens the door to the possibility of experimentally engineering not only UPO, but also AAO enzyme, towards the efficient production of FDCA from HMF.

REFERENCES

[1]   Gallezot, P., Conversion of biomass to selected chemical products. *Chem Soc Rev*, 2012. 41(4): p. 1538-58.

[2]   Huber, G.W., S. Iborra, and A. Corma, Synthesis of transportation fuels from biomass: chemistry, catalysts, and engineering. *Chem Rev*, 2006. 106(9): p. 4044-98.

[3]   Corma, A., S. Iborra, and A. Velty, Chemical routes for the transformation of biomass into chemicals. *Chem Rev*, 2007. 107(6): p. 2411-502.

[4]   U.S. Department of Energy, Energy Information Administration, Independent Statistics & Analysis.(2017, September). *Petroleum Consumption: Transportation and Electric Power Sectors.* Retrieved from https://www.eia.gov/totalenergy/data/monthly/pdf/sec3_21.pdf

[5]   Bozell, J.J. and G.R. Petersen, Technology development for the production of biobased products from biorefinery carbohydrates-the US Department of Energy's "Top 10" revisited. *Green Chem.*, 2010. 12(4): p. 539-554.

[6]   Wierckx, N., *et al.*, *Whole-Cell Biocatalytic Production of 2,5-Furandicarboxylic Acid*, in Microorganisms in Biorefineries, B. Kamm, Editor. 2015, Springer Berlin Heidelberg. p. 207-223.

[7]   Gallezot, P., Process options for converting renewable feedstocks to bioproducts. *Green Chem.*, 2007. 9(4): p. 295-302.

[8]   Werpy, T. and G. Petersen, Top Value Added Chemicals from Biomass: Volume I -- Results of Screening for Potential Candidates from Sugars and Synthesis Gas, in Other Information: PBD: 1 Aug 2004. 2004. p. Medium: ED; Size: 76 pp. pages.

[9]   Jong, E.d., *et al.*, Furandicarboxylic Acid (FDCA), A Versatile Building Block for a Very Interesting Class of Polyesters, in Biobased Monomers, Polymers, and Materials. 2012, *American Chemical Society*. p. 1-13.

[10]   Koopman, F., *et al.*, Efficient whole-cell biotransformation of 5-(hydroxymethyl) furfural into FDCA, 2,5-furandicarboxylic acid. *Bioresour Technol*, 2010. 101(16): p. 6291-6.

[11]   Papageorgiou, G.Z., V. Tsanaktsis, and D.N. Bikiaris, Synthesis of poly(ethylene furandicarboxylate) polyester using monomers derived from renewable resources: thermal behavior comparison with PET and PEN. *Phys Chem Chem Phys*, 2014. 16(17): p. 7946-58.

[12]   Carlini, C., *et al.*, Selective oxidation of 5-hydroxymethyl-2-furaldehyde to furan-2,5-dicarboxaldehyde by catalytic systems based on vanadyl phosphate. *Applied Catalysis A*, General, 2005. 289(2): p. 197-204.

[13]   Van Putten, R.J., *et al.*, Hydroxymethylfurfural, a versatile platform chemical made from renewable resources. *Chem Rev*, 2013. 113(3): p. 1499-597.

[14]   Thomas, S.M., R. DiCosimo, and V. Nagarajan, Biocatalysis: applications and potentials for the chemical industry. *Trends Biotechnol*, 2002. 20(6): p. 238-42.

[15]   Hanke, P.D., Enzymatic oxidation of HMF. 2009, Google Patents.

[16]   Koopman, F., *et al.*, Identification and characterization of the furfural and 5-(hydroxymethyl)furfural degradation pathways of Cupriavidus basilensis HMF14. *Proc Natl Acad Sci* U S A, 2010. 107(11): p. 4919-24.

[17] Dijkman, W.P. and M.W. Fraaije, Discovery and characterization of a 5-hydroxymethylfurfural oxidase from Methylovorus sp. strain MP688. *Appl Environ Microbiol*, 2014. 80(3): p. 1082-90.

[18] Dijkman, W.P., D.E. Groothuis, and M.W. Fraaije, Enzyme-catalyzed oxidation of 5-hydroxymethylfurfural to furan-2,5-dicarboxylic acid. *Angew Chem Int* Ed Engl, 2014. 53(25): p. 6515-8.

[19] Dijkman, W.P., *et al.*, Structure-Based Enzyme Tailoring of 5-Hydroxymethylfurfural Oxidase. *ACS Catal.*, 2015. 5(3): p. 1833-1839.

[20] Carro, J., *et al.*, 5-hydroxymethylfurfural conversion by fungal aryl-alcohol oxidase and unspecific peroxygenase. *FEBS J*, 2015. 282(16): p. 3218-29.

[21] Borrelli, K.W., *et al.*, PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *Journal of Chemical Theory and Computation,* 2005. 1(6): p. 1304-1311.

[22] Piontek, K., *et al.*, Structural basis of substrate conversion in a new aromatic peroxygenase: cytochrome P450 functionality with benefits. *J Biol Chem*, 2013. 288(48): p. 34767-76.

[23] Madhavi Sastry, G., *et al.*, Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*, 2013. 27(3): p. 221-234.

[24] Anandakrishnan, R., B. Aguilar, and A.V. Onufriev, H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modelling and simulations. *Nucleic Acids Research,* 2012. 40(Web Server issue): p. W537-W541.

[25] Jaguar, version 8.4, Schrödinger, LLC, New York, NY, 2014.

[26] QSite, version 6.3, Schrödinger, LLC, New York, NY, 2014.

[27] Kaminski, G.A., *et al.*, Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *The Journal of Physical Chemistry B*, 2001. 105(28): p. 6474-6487.

[28] Bashford, D. and D.A. Case, Generalized born models of macromolecular solvation effects. *Annu Rev Phys Chem*, 2000. 51: p. 129-52.

[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*, 2011. 436: p. 341-350.

[30] Sondergaard, C.R., Olsson, M.H.M, Rostkowski, M. and Jensen, J.H, Improved Treatment of Ligands and Coupling Effects in Empirichemical routes by biochemical catalysis has already shown to lead to potential operating cost reductions (al Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.,* 2011. 7(7): p. 2284-2295.

[31]   Olsson, M.H.M., Sondergaard, C.R., Rostkowski, M., and Jensen, J.H., PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.,* 2011. 7(2): p. 525-537.

[32]   Swain, M., chemicalize.org. *J. Chem. Inf. Model.*, 2012. 52(2): p. 613-615

[33]   Lucas, M.F., *et al*., Simulating Substrate Recognition and Oxidation in Laccases: From Description to Design. *J. Chem. Theory Comput.*, 2017. 13: p. 1462-1467.

[13]   Acebes, S., *et al*., Rational Enzyme Engineering Through Biophysical and Biochemical Modelling. *ACS Catal.,* 2016. 6:p. 1624-1629.

[35]   Babot, E.D., et al., Steroid Hydroxylation by Basidiomycete Peroxygenases: a Combined Experimental and Computational Study. *Appl. Environ. Microbiol.*, 2015. 81 (12):p. 4130-4142.

[36]   Monza, E., *et al.*, Insights into Laccase Engineering from Molecular Simulations: Toward a Binding-Focused Strategy. *J. Phys. Chem. Lett.,* 2015. 6 (8): p. 1447-1453.

[37]   Giacobelli, V.G., *et al.*, Repurposing designed mutants: a valuable strategy for computer-aided laccase engineering - the case of POXA1b. *Catal. Sci. Thecnol.,* 2017. 7: p. 515-523.

[38]   Bendl, J., *et al.* HotSpot Wizard 2.0: automated design of site-specific mutations and smart libraries in protein engineering. *Nucleic Acids Res.*, 2016. 44:p. 479-87.

[39]   P. Schmidtke, *et al.*, MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics*. 2011. 27: p. 3276-85.

[40]   *Desmond Molecular Dynamics System,* version 2.2, D.E. Shaw Research, New York, NY, 2009. Maestro-Desmond Interoperability Tools, version 2.2, Schrödinger, New York.

[41]   Martin, C., *et al.*, Creating a more robust 5-hydroxymethylfurfural oxidase by combining computational predictions with a novel effective library design. *Biotechnol Biofuels*, 2018. 11:56

SUPPLEMENTARY INFORMATION

Towards the efficient production of FDCA: Molecular Modelling and *In Silico* Design of *Aae*UPO, HMFO and AAO enzymes.

Ferran Sancho,[‡a] Marina Cañellas,[‡a] Victor Guallar[ab] and Fátima Lucas[c*]

[a] Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain

[b] ICREA, Passeig Lluis Companys 23, E-08010 Barcelona, Spain

[c*] ZYMVOL BIOMODELLING SL, C/ Almogavers, 165, 08018 Barcelona, Spain; E-mail: flucas@zymvol.com; Tel: +34617927468

[‡] These authors equally contributed to this work

Figure S1. HMF binding in its non-reactive position (hydroxyl group near to heme reactive oxygen). The distance between heme oxygen and its closest HMF hydrogen atom is marked in black. The heme molecule, heme axial residue (Cys36) and protein residues that interact with HMF are colored in grey, and the HMF ligand in green

Figure S2. HMFCA binding at the entrance of the heme cavity. Hydrogen bonds between the ligand and the nearby protein residues are marked in yellow. Heme molecule, heme axial residue (Cys36) and protein residues that interact with the substrate are colored in grey, and the HMFCA ligand in green.
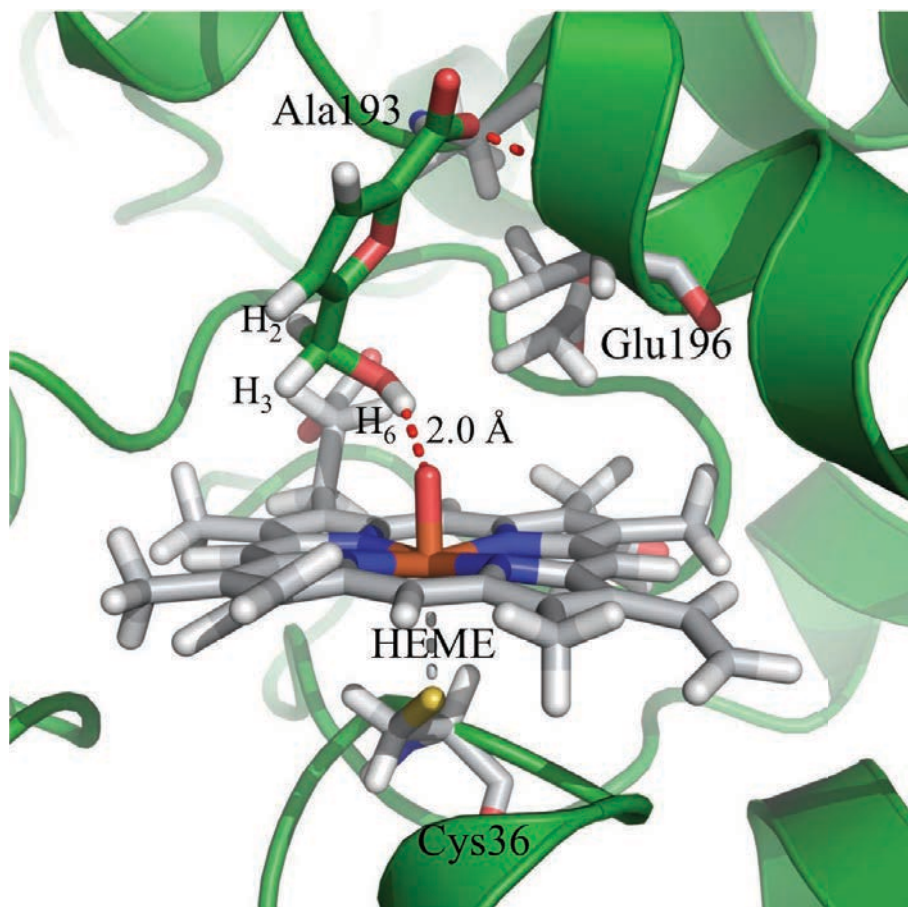
Figure S3. HMFCA binding in its non-reactive position (hydroxyl group forming an hydrogen bond with the heme reactive oxygen). Hydrogen bonds between HMFCA and the nearby protein residues are marked in red, and the distance between the heme oxygen and its closest substrate hydrogen atom is marked in black. Heme molecule, heme axial residue (Cys36) and protein residues that interact with the substrate are colored in grey, and the HMFCA ligand in green.
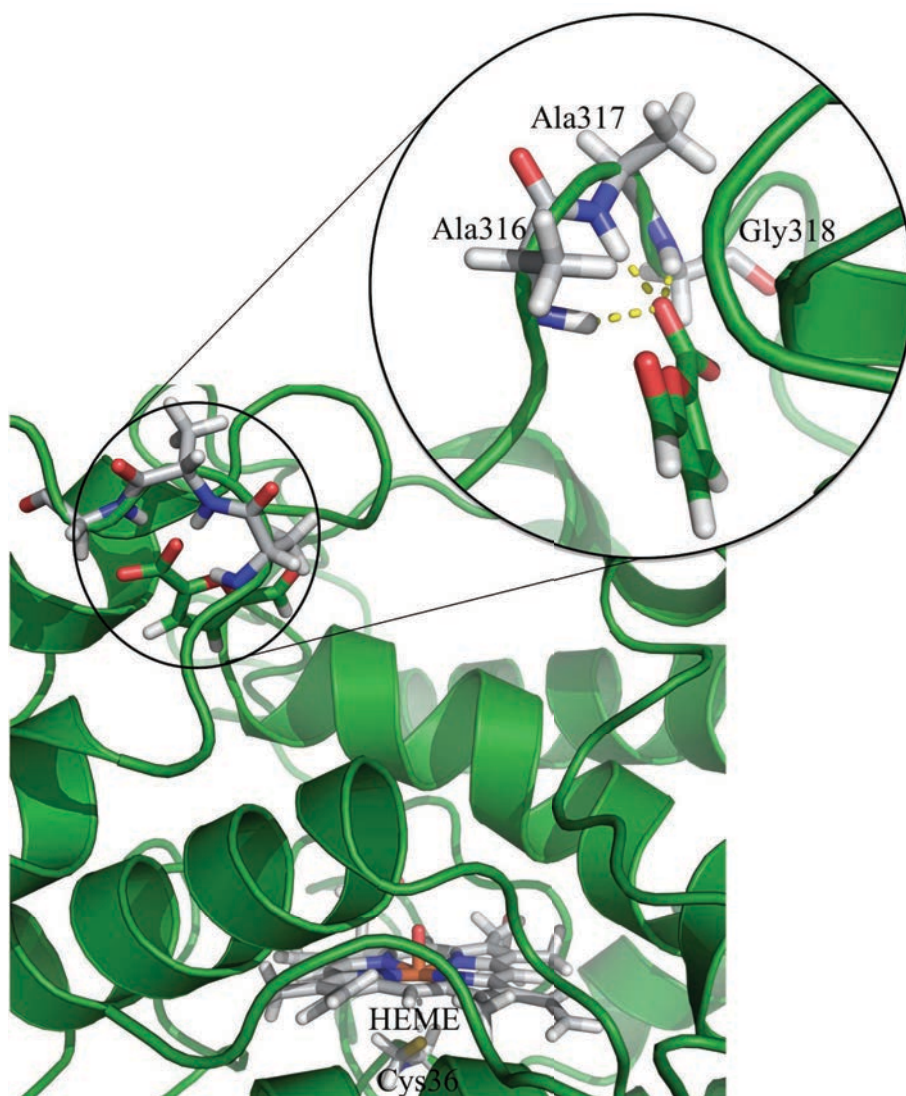
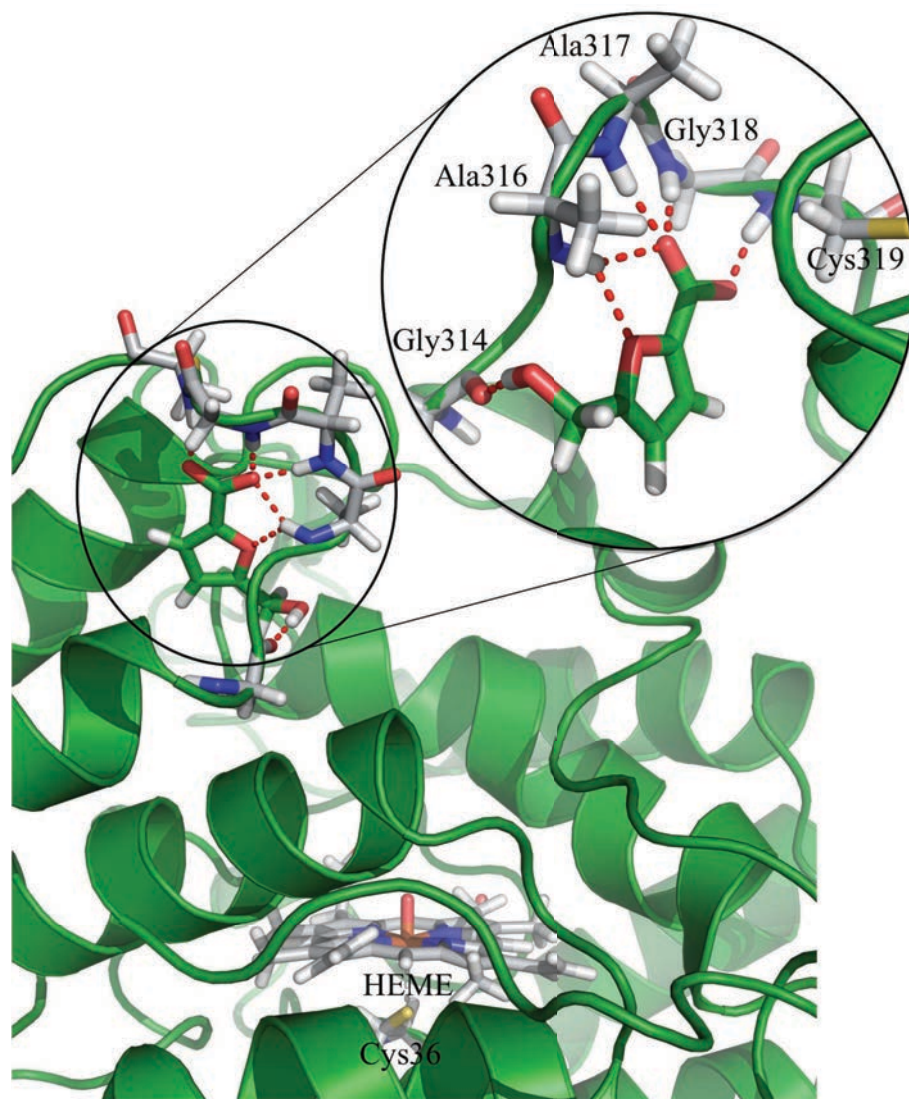**Figure S4.** FFCA binding mode at one of the minima at the entrance of the heme cavity. Hydrogen bonds between the ligand and the nearby protein residues are marked in yellow. Heme molecule, heme axial residue (Cys36) and protein residues that interact with the substrate are colored in grey, and the FFCA ligand in green.
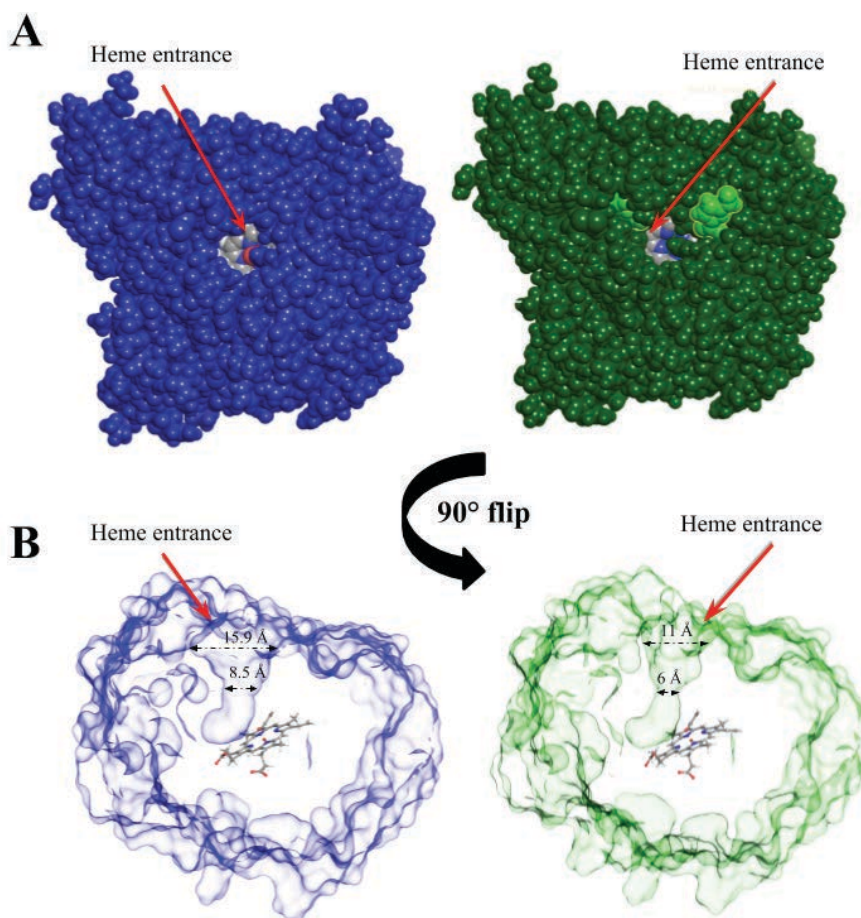
Figure S5. A - Heme entrance frontal view of the UPO protein without FFCA substrate (colored in blue) and with two FFCA ligands placed at the entrance of the heme pocket (colored in green). Protein atoms are displayed with a CPK molecular representation, heme molecule carbon atoms are colored in grey, and FFCA molecules are colored in light green and highlighted in yellow. **B -** Molecular surface representation of the UPO protein without FFCA substrate (blue colored) and with two FFCA molecules placed at the entrance of the heme pocket (green colored). The diameter on two different transversal sections along the heme path is shown in Å.
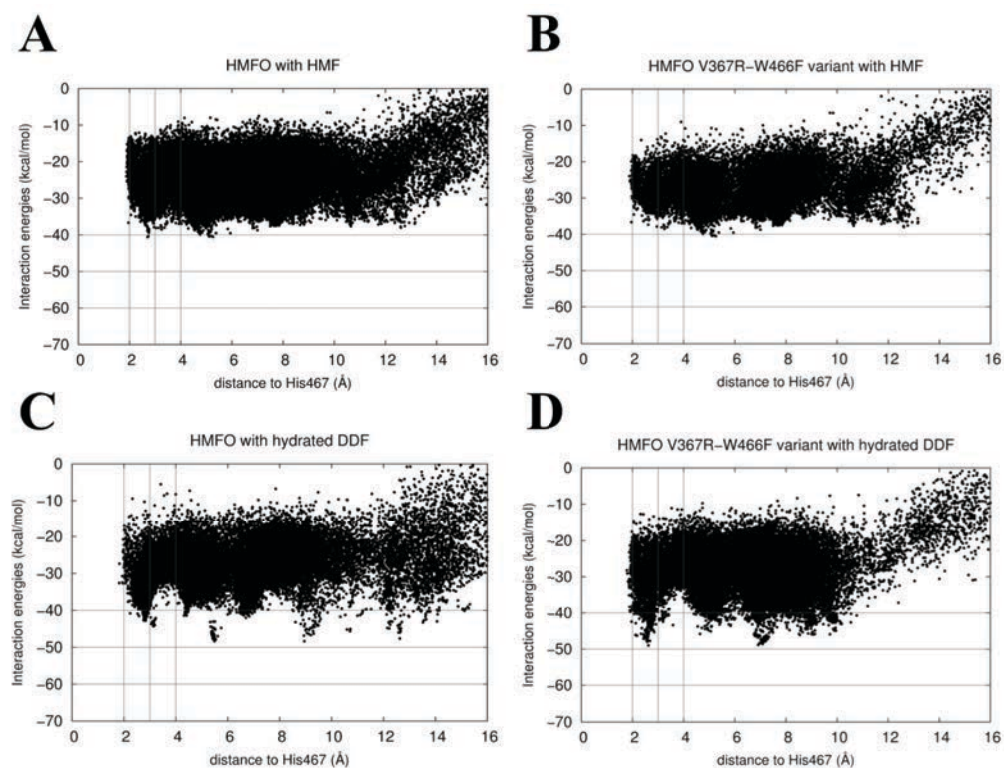
Figure S6. A - Interaction energies vs. ligand distances from PELE simulations in HMFO (left) and HMFO-mut (right). The distances shown (Å) are measured between the reactive N atom in the His467 and the alcohol hydrogen atom from the ligand. The plots correspond to the migration of HMF (A, B) and hydrated DFF (C, D)

Results summary and discussion

4

"What matter most is how well you walk through the fire."

Charles Bukowski

In the previous section (**Chapter 3 - Results**), a very accurate description, both at the computational and experimental level, has been shown for various enzymatic systems and processes. In an attempt to facilitate the comprehension of the work here performed and to avoid falling into false assumptions of authorship, only the results obtained by myself during the development of this thesis are summarized and discussed in the present section. In the cases where the presented results were obtained together with other members of the lab, it will be explicitly specified in the text. Again, results are classified accordingly to the questions that aim to be answered rather than the enzyme system that is studied.

4.1 REACTION MECHANISMS DESCRIPTION

4.1.1 Catalytic surface radical in dye-decolorizing peroxidases

In the present section, the combination of different computational methods at distinct levels of theory is applied to achieve a better understanding of LRET in peroxidases. In this way, the oxidation of reactive blue 19 (RB19) bulky dye by surface residues of *Aau*DyP protein was investigated. First, PELE software was used to describe the protein-ligand recognition process and to identify possible RB19 binding sites in the *Aau*DyP surface. As shown in **Fig. 4.1**, RB19 encounters several favorable positions on the enzyme surface with local minima near to Trp105, Tyr147/Tyr337, Trp207, Tyr285, Trp377 and the heme channel entrance. For an example of PELE substrate exploration see *Article I*, **movie S1**.
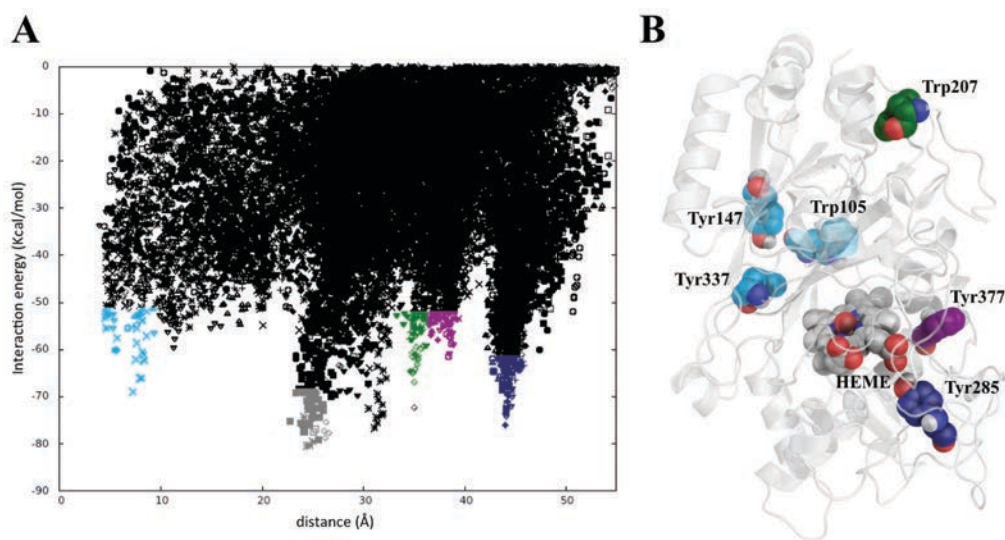

Figure 4.1. (**A**) Local minima identified in the PELE simulations for RB19 exploration on *Aau*DyP surface. The distances shown are between the Tyr147 CZ atom and the ligand's

COM. (**B**) Three-dimensional structure of *Aau*DyP mapping RB19 surface binding sites found by PELE. Residues are colored accordingly to the color of the minima in **A**.

Following the enzymatic process, to clarify if any of the previously identified surface binding sites could indeed oxidize RB19, and to find the LRET pathway from such residue or a set of residues to the heme, we turned into the QM/MM level of theory. First, we assessed which of the identified surface residues had more tendency to be oxidized by performing simple QM/MM pairwise comparisons. Results showed a clear preference for Trp377 residue to be oxidized over others, excepting Trp207 (*Article I*, **Table S2**). However, since electron coupling is known to potentially decay with donor-acceptor distance [154], we can therefore exclude Trp207 as a potential surface oxidant due to its large distance from the heme cofactor. Moreover, in order to establish a preference between the formerly main sites, Trp377 and the one identified by Stritmatter and co-workers (Tyr337) [56], further QM/MM simulations were performed: i) to study residue oxidation by compound I (*Article I*, **Fig. 3A**), and ii) to study RB19 oxidation (*Article I*, **Fig. 3B and C**). Results show Trp377 residue as the more likely to be oxidized and to oxidize the substrate. Additionally, we were able to map the important residues along the LRET pathway from Trp377 to the heme (*Article I*, **Fig. 4**).

Table 4.1. Fraction of electron density present in Trp377 and RB19 in distinct structures along 5 ns MD.

MD snap-shot	RB19 to Trp377	ions	% spin Trp377	% spin RB19	Cl to Trp377	Cl to RB19
1	3.8	Cl	29	28	12.9	3.5
2	3.4	Cl	27	9	14.3	4.5
3	4.2	Cl	36	2	11.1	3.5
4	3.6	Cl	37	14	10.3	3.3
5	5.2	Na/Cl	35	3	10.9	3.2
6	3.6	-	36	0	-	-
7	4.7	Na	22	0	-	-
8	5.0	Cl	31	0	11.0	3.7
9	3.8	Na/Cl	24	0	12.0	4.6
10	3.8	Cl	19	29	13.0	3.2
11	5.0	-	27	0	-	-
12	8.2	Cl	25	0	9.2	9.4
13	7.1	Cl	24	12	6.4	7.6

RB19 to Trp377 corresponds to the distance between RB19 and Trp377 selecting the closest heavy atom. % spin Trp377 is the amount of sity computed in Trp377. % spin RB19 is the amount of spin density computed on RB19. Cl to Trp377 is the distance between the chloride ion and the closest heavy atom of Trp377. Cl to RB19 is the distance between the chloride ion and the closest heavy atom of RB19. All distances are in angstroms.

Furthermore, to study Trp377 and RB19 oxidation dependency on the local electrostatic environment, additional QM/MM calculations were performed on different 5 ns MD structures. All snapshots were different regarding the substrate distance to Trp377 and the electrostatic environment. Results show that the presence of negative chloride ions in the neighbourhood of the substrate is crucial for appreciable oxidation of RB19 (**Table 4.1**), demonstrating the influence of the local electrostatic environment over spin density.

In agreement with spectroscopic and site-directed mutagenesis studies, and in contrast to previously published work [56], our results suggest Trp377 as the main catalytic surface residue responsible for bulky substrate oxidation through a LRET pathway in *Aau*DyP. However, other possible oxidation sites, such as the heme access-channel or a tyrosyl radical, could also have a minor role in this process. In this work, we have demonstrated the importance of using *in silico* methods to obtain an electronic and atomic accurate description of enzymatic processes. Combined with experimental data, our results allowed us to clarify the LRET mechanisms and to explain the role of protein radicals in *Aau*DyP.

4.1.2 Steroid hydroxylation by *Aae*UPO

In *Article II*, the selective hydroxylation of a variety of steroidal compounds (*Article II*, **Fig.1**, compounds A-O) by three different basidiomycete peroxygenases is evaluated experimentally for the first time. Experimental results (*Article II*, **Table 1-4**) show a wide range of hydroxylation yields and selectivities among the different steroids. In an attempt to understand the molecular determinants responsible for the differences in yield and selectivity between the different steroids, a computational study was performed by myself together with F. Lucas and F. Sancho. In this way, substrate diffusion simulations were performed with PELE on a set of 8 representative steroid molecules with *Aae*UPO structure. From the information provided by the energy profiles and trajectories analysis from PELE we were able to rationalize the effect of the structural differences between the substrates and their reactivity in *Aae*UPO.

Our results show that there is not a unique factor that determines the hydroxylation profile of the here studied steroids, but instead it is a result of the sum of many structural determinants including: i) the polar/apolar character of the C_3 substituent, ii) the oxidation/unsaturation degree of the rings, and iii) the nature of the alkyl C_{17} chain. All these factors define the ligand entrance rate by C_{25} end and its positioning at the binding site, and consequently, they have a direct influence in the hydroxylation yield for such compound. In **Table 4.2** a general analysis of the

PELE results for each computationally studied steroid is performed, showing the ligand entrance rate by C_{25} and its minimum distance to the binding site. Further details on energy profiles and interaction maps for each steroid in the heme active site are available in *Article II*.

When talking about regioselectivity, our results show that hydroxylation at positions other than C_{25}, such as C_{24}, C_{26}, C_{27}, and C_{28}, occurs in cases that show a high (energetically favorable) occupation of binding positions near any of these centers. Although hydroxylation of the tertiary C_{25} carbon will always be more favorable [155], hydroxylation can occur at other positions if these present a high fraction of structures optimally positioned. The case of cholesterol's hydroxylation at C_{24}, which has a local minimum on which C_{24} is optimally positioned for reaction, is a clear example (*Article II*, **Fig. S4**). Besides hydroxylation at the side chain, hydroxylated derivatives at the steroid core can be formed if: i) there is a lack of oxygen groups at C_3 position or presence of double bonds in the steroidal ring, and ii) if substrate entrance by C_3 side is favored.

Table 4.2. General PELE trajectories analyses for computationally studied steroidal compounds. (For steroid structure identification refer to *Article II*, **Fig. 1**)

Ligand	% of substrate conversion	% of entrance by C_{25}	Minimum distance to heme (Å)
A (cholesterol)	64	82	2.4
D (sitosterol)	13	39	4.5
F (cholestan-3-one)	42	49	2.8
H (cholesta-3,5-dien-7-	57	58	2.4
J (cholestane)	3	59	3.6
K (cholesta-3,5-diene)	18	45	2.4
M (cholesteryl acetate)	16	35	3.5
O (cholesteryl caprylate)	0	0	5.5

Percentage of substrate conversion obtained within 60 min of reaction, ligand entrance rate by C_{25} side and minimum distance from H_{25} ligand atom to the heme iron (when more than 5 structures are found at this distance) obtained from PELE simulations.

4.1.3 Molecular determinants for selective hydroxylation of vitamins D_2 and D_3 by fungal peroxygenases.

In the present section, atomic level simulations are used to rationalize the differences observed experimentally in cholecalciferol and ergocalciferol's (vitamins D_3 and D_2, respectively) conversion rates and site selectivity by *Aae*UPO and *Cci*UPO (experimental data is shown in *Article III*, **Fig. 2**). Together with F. Lucas, the diffusion of vitamins D_2 and D_3 on the molecular structure of the two UPO enzymes was performed with the PELE software. Analyses of the obtained energy profiles, binding modes and relative distance distributions of the substrates for each protein-ligand system were used to investigate the differences in yield and regioselectivity.
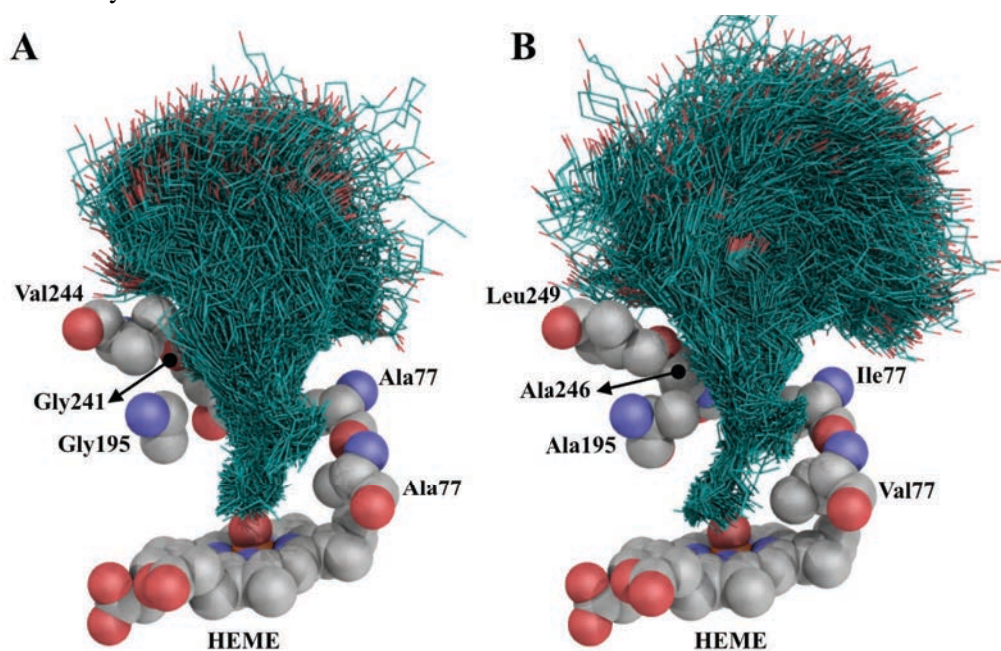
Figure 4.2. Superposition of all the substrate (cholecalciferol) positions along PELE simulations on *Aae*UPO (**A**) and *Cci*UPO (**B**). Heme cofactor and selected residues located on the heme channel are shown as VdW spheres.

In good agreement with experimental conversion rates, PELE results show that *Cci*UPO has the most favorable energy profiles and well-defined minima in the binding site (*Article III*, **Fig. 5**). From a structural point of view, this is caused by a tighter entrance to the heme-binding site in *Cci*UPO, resulting from a higher number of bulky hydrophobic side chains when compared with *Aae*UPO. This smaller cavity leads to a better wrapping of the protein around the substrate at catalytic positions, which is reflected in more favorable energy profiles. In **Fig. 4.2** a comparison between *Aae*UPO and *Cci*UPO is performed highlighting previously

mentioned structural differences between them and their effect on substrate binding. Moreover, our results suggest that the lower reactivity observed for D_2 substrate on both UPO enzymes could be caused by steric clashes between the substrate's extra methyl group at C_{24} and residues at position 69, which hinder a proper substrate positioning (*Article III*, **Fig. S2**).

Additionally, the different site selectivities observed for vitamins D_3 and D_2 hydroxylation by *Aae*UPO and *Cci*UPO were investigated by analyzing the relative distance distribution of the substrates' reactive hydrogen atoms in the active site. In agreement to the strict C_{25} site selectivity observed experimentally for *Cci*UPO, our results show a dominant O-H_{25} frequency distribution at catalytic distances for both compounds in this enzyme. In *Aae*UPO, whereas the fraction of O-H_{25} reactive distances is dominant for D_2, for D_3 the fraction of O-H_{24} reactive distances is considerably higher when compared to the other systems (**Table 4.3**). The decrease on the O-H_{25}/O-H_{24} ratio of favorable structures could explain the observed formation of C_{24} hydroxylation products in cholecalciferol.

Table 4.3. Experimental and computational site selectivity analysis.

Enzymatic system	% C_{25} formation (experimental)	% H_{25} at reactive positions	% H_{24} at reactive positions
D_3 (*Cci*UPO)	100	54.4	5.2
D_3 (*Aae*UPO)	64	27.4	19.6
D_2 (*Cci*UPO)	100	36.2	1.8
D_2 (*Aae*UPO)	100	25.7	1.3

Experimental fraction of C_{25}-hydroxylated product formation, fraction of structures with H_{25} at reactive positions, and fraction of structures with H_{24} at reactive positions for cholecalciferol (D_3) and ergocalciferol (D_3) in *Aae*UPO and *Cci*UPO. Only those PELE structures at distances closer than 3 Å from the heme oxygen atom have been considered for this computational analysis.

4.2 ENHANCED VARIANT RATIONALIZATION

4.2.1 Asymmetric sulfoxidation by engineering the heme pocket of *Aau*DyP

In *Article IV*, the heme pocket of *Aau*DyP was engineered by our collaborators with the aim of providing it with new catalytic activities of biotechnological interest. In this way, while the native protein was unable to perform sulfoxidation

reactions, two individual mutations at the heme-binding site provided sulfoxidation activity on methyl-phenyl sulfide (MPS) and methyl-*p*-tolyl sulfide (MTS) substrates. Interestingly, F359G variant showed strict *s* stereoselectivity while L357G variant yielded to a racemic mixture of both isomers (*Article IV*, **Fig. 3**). Here, with the use of molecular modelling techniques at different levels of theory, we were able to rationalize the differences in yield and selectivity observed during substrate sulfoxidation by the native *Aau*DyP and its two variants.
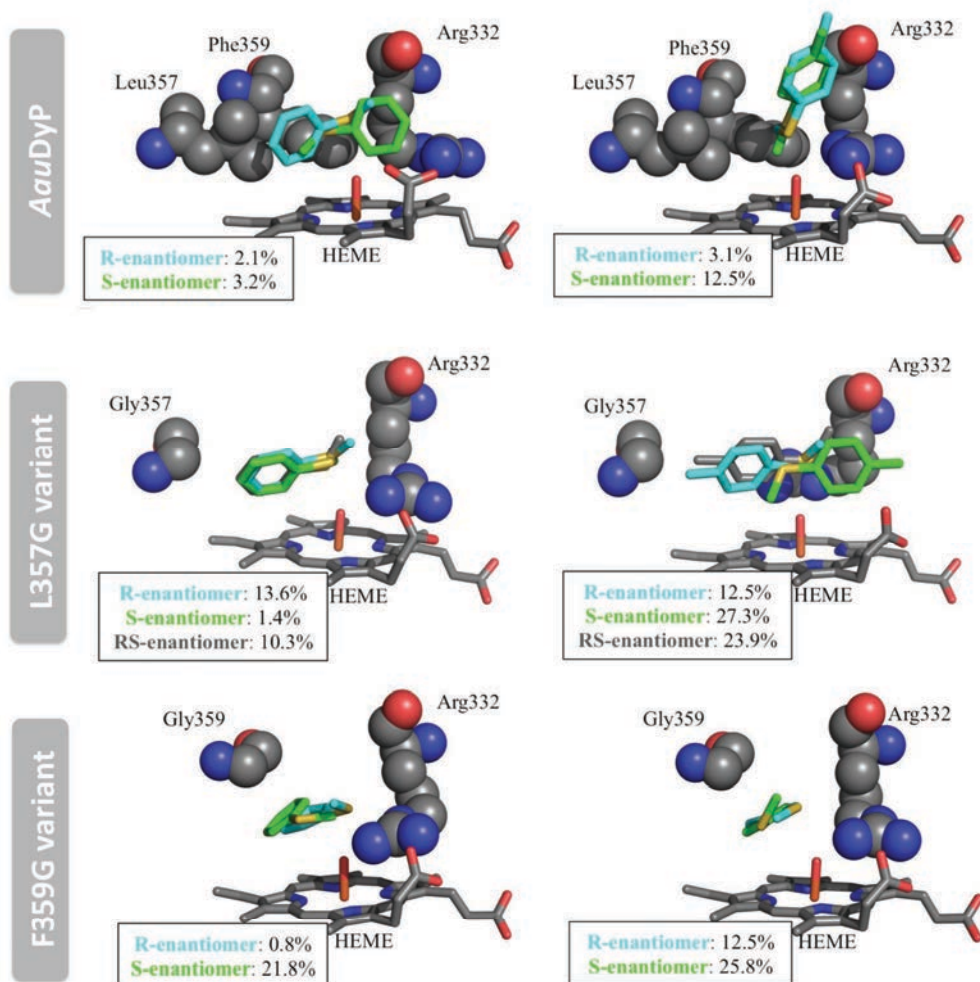

Figure 4.3. Substrate positioning at the heme pocket for pro-*R*, pro-*S*, and pro-*R/S* MPS (left) and MTS (right) on *Aau*DyP, L357G and F359G variants. Pro-*S*, pro-*R* and pro-*S/R* substrate positions are shown as CPK sticks and colored in green, cyan and gray, respectively. The spin density population on the substrate, obtained from QM/MM calculations, is indicated in each image.

First, PELE substrate diffusions were performed on the three DyP structures. In agreement with the undetectable (on MTS) or very low (on MPS) activity shown experimentally for *Aau*DyP, PELE results showed a more favorable protein-substrate catalytic complex (higher number of structures below 4 Å with better interaction energies) for the two variants when compared to the native DyP (*Article IV*, **Fig. 4**). Interestingly, a higher number of trajectories are seen at catalytic distances for those conditions where over 90% conversion in 30 min is observed experimentally: F359G reaction with MTS and MPS, and L357G reaction with MPS. Moreover, to investigate the differences on stereoselectivity observed between the two variants, a study of the pro-*R* and pro-*S* positioning tendency for MPS and MTS was performed with PELE structures. As shown in **Fig. 4.3**, while the presence of bulky residues (Arg332, Leu357 and Phe359) in the heme pocket hinders the proper substrate positioning in *Aau*DyP, in DyP variants we find an easier substrate ring positioning. However, contrary to the experimentally observed selectivity, substrates on both variants show a preference to adopt pro-*R* positions.
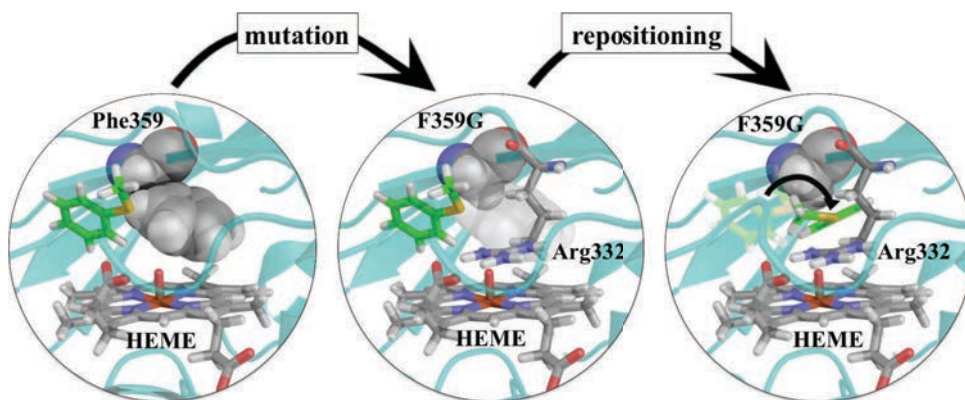
Figure 4.4. F359G mutation effect on pro-*R* MPS substrate positioning with respect Arg332.

In an attempt to clarify this apparent contradiction with the experimental results, we turned into QM/MM methodology and compute substrate spin densities on the best substrate positioned structures. In agreement with experimental results, QM/MM calculations showed: i) a significant larger spin density in DyP variants when compared to *Aau*DyP; and ii) a substantial difference between pro-*R* and pro-*S* structures spin density in F359G variant (**Fig. 4.3**). Our results suggest that in F357G variant, even though a higher fraction of structures are placed favoring the formation of an *R* enantiomer, pro-*S* structures are easily oxidized, which explain the *S* stereoselectivity experimentally observed for this variant. From a structural point of view, we observed that these changes in spin density are correlated with the position adopted by the substrate with respect Arg332. In this way, F357G

mutation induces a change in the ligand positioning that leads to a closer interaction of pro-*R* structures with this arginine residue, which negatively affects substrate oxidation (**Fig. 4.4** and *Article IV*, **Fig. 7**).

4.2.2 Synthesis of 1-naphthol by an *Aae*UPO variant

Our collaborators showed possible to experimentally improve *Aae*UPO towards the efficient selective production of 1-naphthol. To do so, peroxygenase activity (*P*; O-transferring two-electron oxidations) has to be improved or protected while peroxidative activity (*p*; one-electron oxidations), which favors unwanted metabolite production, has to be reduced. In *Article V*, a new *Aae*UPO variant was produced by taking as starting point PaDa-I variant (previously created by DE [64]) after two rounds of DE. The final variant, JaWa, harbored two additional mutations (G241D and R257K) and showed a 2-fold increase on peroxygenase activity (screened with naphthalene) and half the peroxidative activity (screened with 2,6-dimethylzenol (DMP)) (*Article V*, **Table 2**). In this work, to improve our understanding of the atomic mechanistic details responsible for the observed differences between PaDa-I and JaWa variants, we turned into molecular modelling.

First, to explain the effect of the two introduced mutations on the peroxygenase activity, naphthalene exploration at the heme site of PaDa-I and JaWa was performed with PELE software. Results show that in the JaWa variant G241D mutation induces several conformational changes that lead to a redistribution of the binding site occupation, and more importantly, to the displacement of the α-helix hosting the catalytic Arg189-Glu196 pair (*Article V*, **Fig. 5**). As shown in *Article V*, **Fig. 6A**, the rearrangement of the JaWa binding site induces a better naphthalene positioning in the heme-binding pocket, in agreement with its improved K_m when compared to PaDa-I. Additionally, G241D-induced structural changes affect not only naphthalene's binding but also its hydroxylation. In this way, the above-mentioned α-helix changes induce a new alternative position for Arg189 in JaWa variant, closer to the substrate (*Article V,* **Fig. 6B**). In agreement with the lower k_{cat} value observed experimentally for JaWa, this change in the substrate's electrostatic environment increases by 1 kcal.mol^{-1} the naphthalene's ionization energy in JaWa, leading to a higher activation barrier (computed by QM/MM calculations). For a visual comparison of the above-mentioned structural differences between JaWa and PaDa-I variants and its effect on naphthalene's binding, refer to *Article V*, **movie S1**.

On the other hand, changes on peroxidative activity were also investigated. Following the protocol previously employed in *Article I* to identify catalytic surface residues in DyP, PELE global DMP exploration was performed to identify the most favorable minima. Further QM/MM calculations on potential oxidation sites identified Trp24 surface residue to be the most favorable site for peroxidative activity (*Article V*, **Fig. S5A**), in agreement with site-directed mutagenesis experiments. Additionally, further PELE simulations with DMP substrate showed an additional access to the heme site for this compound (**Fig. 4.5**), very close to 257 position, suggesting that R257K mutation might alter both DMP' k_{cat} and K_m.
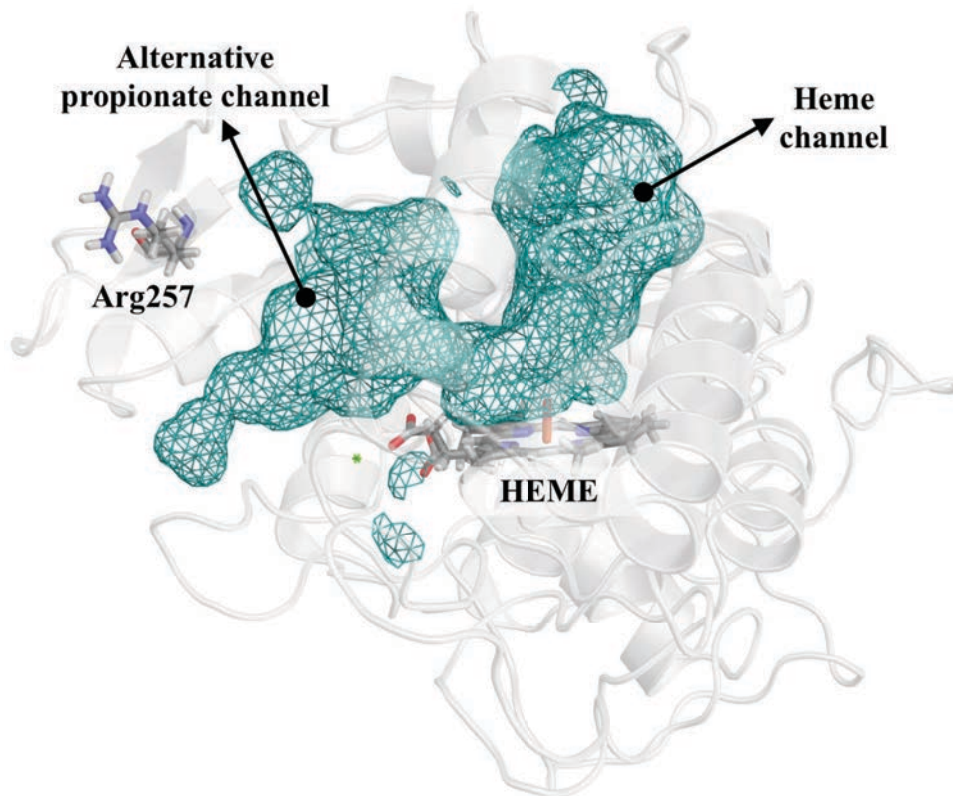
Figure 4.5. *Aae*UPO structure showing the alternative access to the heme site, placed nearby position 257, found for DMP substrate. The heme and the alternative propionate access channels are represented as a mesh surface.

4.2.3 Selective synthesis of 5'-OHP by an evolved *Aae*UPO

As already mentioned, the enhancement of the *P:p* ratio is fundamental to perform C-H oxyfunctionalization with UPO enzymes. This is of especial importance when hydroxylating aromatic compounds, such as propranolol, since the peroxygenative activity products become substrates of UPO's peroxidative

activity, which ultimately leads to undesired phenoxyl radical coupling that can only be circumvented by the use of expensive radical scavengers (*Article VI*, **Scheme 1**). In *Article VI*, our collaborators tailored *Aae*UPO towards the high-efficient selective synthesis of 5'-hydroxypropranolol (5'-OHP) from propranolol in the absence of radical scavengers. The resulting variant (named SoLo), which was designed by taking JaWa variant as starting point, harbored only an additional mutation (F191S), and was found to enhance the *P:p* ratio by 1.8-fold. In this work, to study the differences found experimentally between the *Aae*UPO, PaDa-I, JaWa and SoLo variants propranolol hydroxylation (peroxygenative activity) and ABTS oxidation (peroxidative activity) (refer to *Article VI*, **Table 1**, for a biochemical characterization of all the variants), we turned into molecular modelling.

First, to study the effect of mutation on propranolol hydroxylation, we modeled propranolol diffusion to the active site on each variant. As shown in *Article VI*, **Fig. 3**, PELE resulting binding energies and distances correlate largely with experimental K_m. In this way, when analyzing the relative distance distribution of the propranolol' reactive carbon atom (C_5) in the active site, we find a significantly higher number of optimal poses for those variants with better K_m (JaWa and SoLo). Simulations showed that the better K_m observed for JaWa and SoLo variants arise from G241D mutation, which seems to anchor the substrate in a favorable position for hydroxylation (*Article VI*, **movie S1**). Moreover, additional MD simulations and analyses of the resulting trajectories showed that F191S causes an enlargement of the heme pocket in SoLo variant. These structural variations lead to a modest destabilization of the propranolol catalytic orientation in SoLo when compared to JaWa variant, in agreement with the slightly better observed K_m for this last variant (*Article VI*, **Fig. 4**). In **Fig. 4.6**, a comparison between propranolol's binding site positioning among the four different UPO proteins is shown, emphasizing the role of G241D and F191S mutations on substrate positioning.

On the other hand, the effect of mutation on the peroxidative activity was also investigated with PELE by modelling 5'-OHP diffusion to JaWa and SoLo active sites. Results show a significantly smaller concentration of catalytic 5'-OHP positions, as well as shorter substrate' residence times in the binding site, in SoLo when compared to JaWa (*Article VI*, **Fig. S8**). Taken together, computational results show that while F191S mutation is the main responsible for the decrease in peroxidative activity in SoLo, G241D mutation is the main driver of the improved peroxygenative activity.
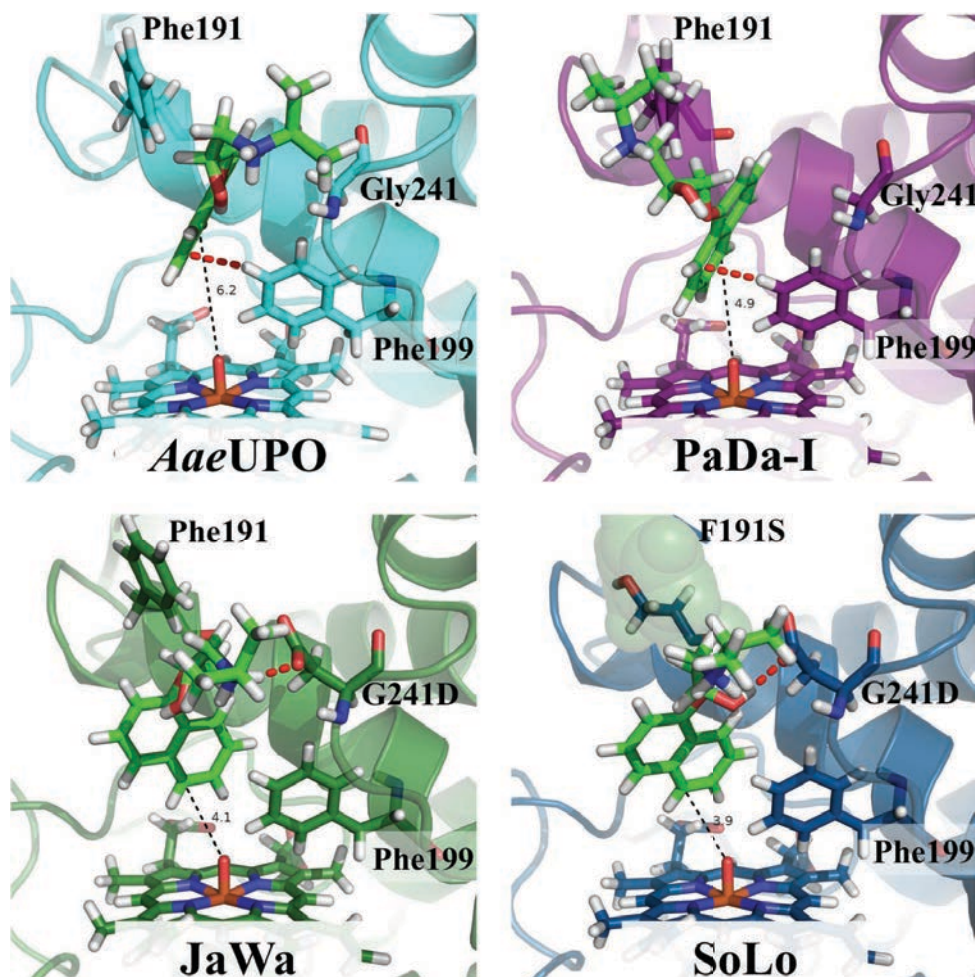
Figure 4.6. Propranolol binding mode comparison between *Aae*UPO, PaDa-I, JaWa and SoLo enzymes. Critical hydrogen bonds between the ligand (colored in green) and nearby protein residues are marked in red, and the distances (in Å) between the reactive propranolol' C_5 atom and the heme oxygen are colored in black.

4.3 COMPUTATIONAL ENZYME REDESIGN

4.3.1 Towards the efficient production of FDCA: Molecular modelling and *in silico* design of *Aae*UPO enzyme

In *Article VII*, together with F. Sancho, an extensive computational study was performed to investigate the transformation of 5-hydroxymethylfurfural (HMF) to 2,5-furandicarboxylic acid (FDCA) by *Aae*UPO, AAO, and HMFO

enzymes. However, since the work performed in this thesis is centered in the study of *Aae*UPO, I will uniquely focus on this last enzymatic system.

First, to investigate at the molecular level the details that govern the enzymatic transformation of HMF to FDCA, substrate migration simulations were performed for each step of the reaction (*Article VII*, A-C reaction in **Fig. 1**) using PELE. From the simulations, energy profiles and binding modes for HMF, 2,5-hydroxymethylfurancarboxylic acid (HMFCA) and 2,5-formylfurancarboxylic acid (FFCA) substrates on *Aae*UPO were obtained (*Article VII*, **Fig 2**), enabling us to rationalize the experimental yields obtained for each different substrate by Carro *et al.* [67]. For HMF, results show that the ligand can be placed in the UPO binding site in two different minima. Whereas in the first minima the substrate is optimally positioned for reaction (*Article VII*, **Fig. 2A**, cyan colored), in the second one the HMF reactive hydrogen is unable to approach the O-Heme to a reactive position (*Article VII*, **Fig. 2A**, green colored). The presence of this unproductive competing minimum could be the reason behind the slow HMFCA production observed experimentally.

On the other hand, simulations with HMFCA and FFCA substrates show very different energy landscapes compared to HMF, where strong minima are located at the heme channel entrance (*Article VII*, **Fig. 2B and C**, orange colored). Interestingly, additional simulations showed that these energy minima lead to a reduction of the diameter of the heme channel entrance (*Article VII*, **Fig. S5**). Thus, although HMFCA and FFCA substrates can arrive to the heme-binding site and adopt catalytic positions (*Article VII*, **Fig. 2B and C**, cyan colored), this narrowed entrance, along with the consequent difficulty of the ligands to reach the active site, explain the slower catalysis shown for HMFCA and FFCA substrates when compared to HMF. Moreover, simulations with FFCA show that although the substrate is able to a reach to the binding site, the distances between the O-Heme and the reactive hydrogen atom are relatively large. This nonoptimal distance, along with the difficulty of the substrate to reach the binding site, can explain why FFCA reaction is the limiting step of the whole enzymatic transformation.

Besides, valuable molecular-based information extracted from PELE simulations, along with an *in silico* enzyme evolution protocol, were synergistically used to model *Aae*UPO towards the more efficient production of FDCA from FFCA in an attempt to theoretically enhance the whole reaction. To do so, two main problems in FFCA conversion had to be overcome: i) the presence of global energy minima at the entrance of the heme path that hinders FFCA entrance, and ii) the non-optimal substrate positioning in the heme-binding site. To overcome the

first problem, T242A and A317P mutations were proposed by rational design to reduce these minima. Additionally, an *in silico* evolution protocol was applied to *Aae*UPO to improve FFCA binding mode. Best resulting variants were: UPO-*mut1*, which hosted mutations A77S and A193S apart from those proposed by rational design; and UPO-*mut2*, which added F121K mutation to T242A and A327P. To computationally characterize the obtained variants, PELE FFCA diffusions were performed showing an improvement of the energy landscape and the substrate binding upon mutation in both UPO-*mut1* (*Article VII*, **Fig. 3**) and UPO-*mut2* (**Fig. 4.7**). Finally, it is worth mentioning that the two computationally designed UPO variants are currently being experimentally characterized by our collaborators, and that we expect to know their kinetic parameters very soon.
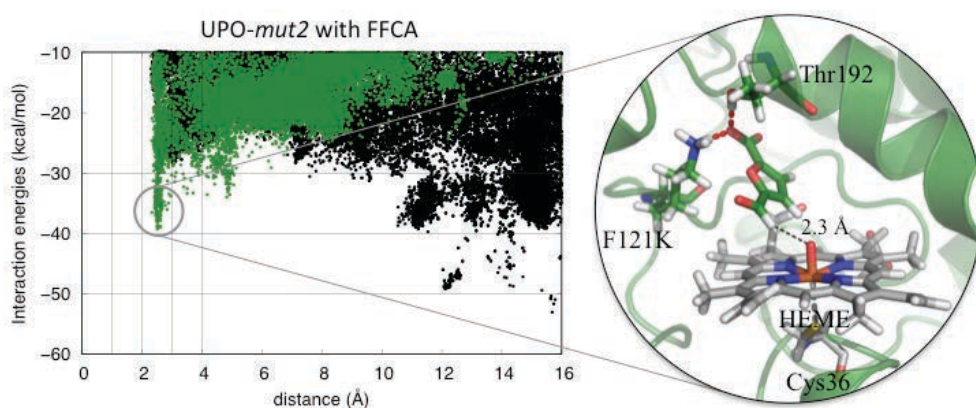
Figure 4.7. UPO-*mut2* variant with FFCA substrate. On the left there is a representation of the interaction energy of each PELE structure against its reactive distance for the *Aae*UPO (colored in black) and for the UPO-*mut2* (colored in green). Reactive distance (Å) is measured between the H_2O_2-activated heme (compound I) and the substrate reactive aldehyde hydrogen. On the right, a representative structure for the specified minimum is shown.

Conclusions

5

"Most human beings have an almost infinite capacity for taking things for granted."

Aldous Huxley, *Brave New World*

The general conclusion that we can derive from the thesis here presented is that molecular modelling techniques, at different levels of theory, are essential tools for the study and design of enzymatic systems, in particular peroxidases, at a reduced cost. Our results have demonstrated the importance of using *in silico* methods to get atomic and electronic insights into processes involved in enzyme catalysis, not only in natural occurring enzymes but also in experimentally evolved variants. Moreover, the better understanding of the molecular principles under enzyme activity, along with the use of *in silico* semi-rational redesign methods, has proven to be an efficient protocol for tailoring enzymes towards desired properties. Besides, the specific conclusions of this thesis are:

i) **Reaction mechanism description.**

- A computational methodology that combines PELE global substrate exploration and QM/MM calculations was successfully used to identify catalytic surface residues responsible for bulky substrate oxidations.

- LRET pathway from Trp377 surface residue to the heme was identified by QM/MM e-Pathway methodology in *Aau*DyP.

- PELE substrate diffusion in UPO enzymes' binding site, along with the subsequent analysis of the resulting energy profiles and substrate binding modes, have enabled us to understand the atomic details under experimentally observed differences in yield and selectivity among distinct enzymatic systems.

- Distance and angle distribution analysis of binding site structures coming from PELE simulations has proven to be an efficient tool for rationalizing experimentally observed regioselectivity differences in peroxidases.

- For *Aau*UPO, a direct correlation between the experimental conversion rate and the steroids entrance rate by C_{25} end (from PELE simulations) could be established.

ii) **Enhanced variant rationalization.**

- The combined use of PELE and QM/MM simulations has been successfully used to provide a rational explanation at the atomic level of the differences, both in yield and selectivity, observed experimentally by DyP enzyme and its variants.

- Molecular modelling techniques have enabled us to understand how mutation-induced structural changes have a repercussion in both substrate' K_M and k_{cat}.

- Changes in the catalytic acid-base pair positioning upon mutation, which alter the binding site electrostatic environment, seem to have a direct effect on substrate hydroxylation/sulfoxidation in peroxidases.

- Analyses of the PELE energy landscapes, binding distances, relative distance distributions and residence times in the binding site showed to be a valuable tool to rationalize the observed differences in peroxidative and peroxygenative activities among different UPO variants.

iii) **Computational enzyme redesign.**

- PELE simulations were successfully used to dissect the reasons behind the *Aae*UPO slow FDCA production.

- New UPO variants were designed for an efficient production of FDCA by synergistically using valuable molecular-based information extracted from PELE simulations along with an *in silico* enzyme evolution protocol.

References

"Pies, ¿para qué los tengo si tengo alas pa' volar?"

"Feet, why do I want them if I have wings to fly?"

Frida Kahlo

1. Copeland, R.A., *Enzymes : a practical introduction to structure, mechanism, and data analysis*. 2nd ed. ed. 2000, New York ; Chichester: J. Wiley.

2. Cooper, G., *The cell: A molecular approach.* The Central Role of Enzymes as Biological Catalysts. 2000, Sunderlan, United States of America: Sinauer Associates.

3. Zaborsky, O.R., *9 - Enzymes: Biological Catalysts*, in *Advanced Materials in Catalysis*. 1977, Academic Press. p. 267-291.

4. Yoo, Y.J.a., et al., *Fundamentals of enzyme engineering*.

5. Messing, R.A., *Immobilized enzymes for industrial reactors*. 1975, New York ; London: Academic Press.

6. Kirk, O., T.V. Borchert, and C.C. Fuglsang, *Industrial enzyme applications.* Current Opinion in Biotechnology, 2002. **13**(4): p. 345-351.

7. Traut, T., *Allosteric Regulatory Enzymes*. 2008, Springer: Boston, MA, Unites States of America.

8. Payen, A. and J.F. Persoz, *Annales de chimie et de physique*, in *"Mémoire sur la diastase, les principaux produits de ses réactions et leurs applications aux arts industriels "(Memoir on diastase, the principal products of its reactions, and their applications to the industrial arts)*. 1833, Chez Crochard: Paris. p. 73-92.

9. Khan, M.Y. and F. Khan, *Principles of Enzyme Technology*. 2015, Delhi, India: PHI Learning.

10. Laidler, K.J. and M.C. King, *Development of transition-state theory.* The Journal of Physical Chemistry, 1983. **87**(15): p. 2657-2664.

11. Murray, H.C. and D.H. Peterson, *Oxygenation of steroids by mucorales fungi*. 1952, Google Patents.

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

13. SMYTH, D.G., W.H. STEIN, and S. MOORE, *The sequence of amino acid residues in bovine pancreatic ribonuclease: revisions and confirmations.* J Biol Chem, 1963. **238**: p. 227-34.

14. Monod, J., J. Wyman, and J.-P. Changeux, *On the nature of allosteric transitions: A plausible model.* Journal of Molecular Biology, 1965. **12**(1): p. 88-118.

15. Visuri, K. and A.M. Klibanov, *Enzymatic production of high fructose corn syrup (HFCS) containing 55% fructose in aqueous ethanol.* Biotechnology and Bioengineering, 1987. **30**(7): p. 917-920.

16. Hummel, W. and M.-R. Kula, *Dehydrogenases for the synthesis of chiral compounds.* European Journal of Biochemistry, 1989. **184**(1): p. 1-13.

17. Robinson, P.K., *Enzymes: principles and biotechnological applications.* Essays In Biochemistry, 2015. **59**: p. 1.

18. Khan, F.A., *Biotechnology Fundamentals.* Proteins and Proteomics. 2015: CRC press.

19. O'Connor, S.M. and J.U. Adams, *Essentials of Cell Biology.* 2010, NPG Education: Cambridge, MA.

20. Shoda, S.-i., et al., *Enzymes as Green Catalysts for Precision Macromolecular Synthesis.* Chemical Reviews, 2016. **116**(4): p. 2307-2413.

21. Hagen, J., *Biocatalysis*, in *Industrial Catalysis.* 2015, Wiley-VCH Verlag GmbH & Co. KGaA. p. 81-98.

22. Bar-Even, A., et al., *The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters.* Biochemistry, 2011. **50**(21): p. 4402-4410.

23. Jemli, S., et al., *Biocatalysts: application and engineering for industrial purposes.* Critical Reviews in Biotechnology, 2016. **36**(2): p. 246-258.

24. Goswami, A. and J.D. Stewart, *Chapter 1 - Introduction, Types of Reactions, and Sources of Biocatalysts*, in *Organic Synthesis Using Biocatalysis.* 2016, Academic Press. p. 1-38.

25. Barrozo, A., et al., *Computational Protein Engineering: Bridging the Gap between Rational Design and Laboratory Evolution.* International Journal of Molecular Sciences, 2012. **13**(10).

26. Monza, E., et al., *Molecular Modeling in Enzyme Design, Toward In Silico Guided Directed Evolution*, in *Directed Enzyme Evolution: Advances and Applications*, M. Alcalde, Editor. 2017, Springer International Publishing: Cham. p. 257-284.

27. Li, S., et al., *Technology Prospecting on Enzymes: Application, Marketing and Engineering.* Computational and Structural Biotechnology Journal, 2012. **2**: p. e201209017.

28. Garske, A.L., G. Kapp, and J.C. McAuliffe, *Industrial Enzymes and Biocatalysis*, in *Handbook of Industrial Chemistry and Biotechnology*, J.A. Kent, T.V. Bommaraju, and S.D. Barnicki, Editors. 2017, Springer International Publishing: Cham. p. 1571-1638.

29. Singh, K.R., et al., *From Protein Engineering to Immobilization: Promising Strategies for the Upgrade of Industrial Enzymes.* International Journal of Molecular Sciences, 2013. **14**(1).

30. Hofrichter, M. and R. Ullrich, *New Trends in Fungal Biooxidation*, in *Industrial Applications*, M. Hofrichter, Editor. 2011, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 425-449.

31. Zámocký, M., et al., *Independent evolution of four heme peroxidase superfamilies.* Archives of Biochemistry and Biophysics, 2015. **574**(Supplement C): p. 108-119.

32. Hofrichter, M., et al., *New and classic families of secreted fungal heme peroxidases.* Applied Microbiology and Biotechnology, 2010. **87**(3): p. 871-897.

33. Vlasits, J., et al., *Mechanisms of catalase activity of heme peroxidases.* Archives of Biochemistry and Biophysics, 2010. **500**(1): p. 74-81.

34. de Montellano, P.R.O., *Catalytic Mechanisms of Heme Peroxidases*, in *Biocatalysis Based on Heme Peroxidases: Peroxidases as Potential Industrial Biocatalysts*, E. Torres and M. Ayala, Editors. 2010, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 79-107.

35. Zámocký, M., et al., *Turning points in the evolution of peroxidase–catalase superfamily: molecular phylogeny of hybrid heme peroxidases.* Cellular and Molecular Life Sciences, 2014. **71**(23): p. 4681-4696.

36. Fawal, N., et al., *PeroxiBase: a database for large-scale evolutionary analysis of peroxidases.* Nucleic Acids Research, 2013. **41**(D1): p. D441-D444.

37. Poulos, T.L., *Heme Enzyme Structure and Function.* Chemical reviews, 2014. **114**(7): p. 3919-3962.

38. Ruiz-Dueñas, F.J. and A.T. Martínez, *Structural and Functional Features of Peroxidases with a Potential as Industrial Biocatalysts*, in *Biocatalysis Based on Heme Peroxidases: Peroxidases as Potential Industrial Biocatalysts*, E. Torres and M. Ayala, Editors. 2010, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 37-59.

39. Yoshioka, S., et al., *Roles of the axial push effect in cytochrome P450cam studied with the site-directed mutagenesis at the heme proximal site.* Journal of Inorganic Biochemistry, 2000. **81**(3): p. 141-151.

40. Green, M.T., *Imidazole-Ligated Compound I Intermediates: The Effects of Hydrogen Bonding.* Journal of the American Chemical Society, 2000. **122**(39): p. 9495-9499.

41. de Ropp, J.S., et al., *Influence of the Distal His in Imparting Imidazolate Character to the Proximal His in Heme Peroxidase: 1H NMR Spectroscopic Study of Cyanide-Inhibited His42→Ala Horseradish*

Peroxidase. Journal of the American Chemical Society, 2002. **124**(37): p. 11029-11037.

42.   Morozov, A.N. and D.C. Chatfield, *How the Proximal Pocket May Influence the Enantiospecificities of Chloroperoxidase-Catalyzed Epoxidations of Olefins.* International Journal of Molecular Sciences, 2016. **17**(8): p. 1297.

43.   Groves, J.T., *Using push to get pull.* Nature Chemistry, 2014. **6**: p. 89.

44.   Singh, R., et al., *Distal Heme Pocket Residues of B-type Dye-decolorizing Peroxidase: ARGININE BUT NOT ASPARTATE IS ESSENTIAL FOR PEROXIDASE ACTIVITY.* Journal of Biological Chemistry, 2012. **287**(13): p. 10623-10630.

45.   Carlini, C., et al., *Selective oxidation of 5-hydroxymethyl-2-furaldehyde to furan-2,5-dicarboxaldehyde by catalytic systems based on vanadyl phosphate.* Applied Catalysis A, General, 2005. **289**(2): p. 197-204.

46.   Poulos, T.L. and J. Kraut, *The stereochemistry of peroxidase catalysis.* Journal of Biological Chemistry, 1980. **255**(17): p. 8199-8205.

47.   Yarman, A., et al., *Can peroxygenase and microperoxidase substitute cytochrome P450 in biosensors.* Bioanalytical Reviews, 2011. **3**(2): p. 67-94.

48.   Bormann, S., et al., *Specific oxyfunctionalisations catalysed by peroxygenases: opportunities, challenges and solutions.* Catalysis Science & Technology, 2015. **5**(4): p. 2038-2052.

49.   Ullrich, R. and M. Hofrichter, *Enzymatic hydroxylation of aromatic compounds.* Cellular and Molecular Life Sciences, 2007. **64**(3): p. 271-293.

50.   Yoshida, T. and Y. Sugano, *A structural and functional perspective of DyP-type peroxidase family.* Archives of Biochemistry and Biophysics, 2015. **574**(Supplement C): p. 49-55.

51.   Strittmatter, E., et al., *First Crystal Structure of a Fungal High-redox Potential Dye-decolorizing Peroxidase: SUBSTRATE INTERACTION SITES AND LONG-RANGE ELECTRON TRANSFER.* Journal of Biological Chemistry, 2013. **288**(6): p. 4095-4102.

52.   Colpa, D.I., M.W. Fraaije, and E. van Bloois, *DyP-type peroxidases: a promising and versatile class of enzymes.* Journal of Industrial Microbiology & Biotechnology, 2014. **41**(1): p. 1-7.

53.   Pühse, M., et al., *Marasmius scorodonius extracellular dimeric peroxidase — Exploring its temperature and pressure stability.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2009. **1794**(7): p. 1091-1098.

54. Strittmatter, E., et al., *The toolbox of Auricularia auricula-judae dye-decolorizing peroxidase – Identification of three new potential substrate-interaction sites.* Archives of Biochemistry and Biophysics, 2015. **574**(Supplement C): p. 75-85.

55. Yoshida, T., et al., *The catalytic mechanism of dye-decolorizing peroxidase DyP may require the swinging movement of an aspartic acid residue.* FEBS Journal, 2011. **278**(13): p. 2387-2394.

56. Strittmatter, E., et al., *Radical formation on a conserved tyrosine residue is crucial for DyP activity.* Archives of Biochemistry and Biophysics, 2013. **537**(2): p. 161-167.

57. Baratto, M.C., et al., *Redox-Active Sites in Auricularia auricula-judae Dye-Decolorizing Peroxidase and Several Directed Variants: A Multifrequency EPR Study.* The Journal of Physical Chemistry B, 2015. **119**(43): p. 13583-13592.

58. Ullrich, R., et al., *Novel Haloperoxidase from the Agaric Basidiomycete Agrocybe aegerita Oxidizes Aryl Alcohols and Aldehydes.* Applied and Environmental Microbiology, 2004. **70**(8): p. 4575-4581.

59. Molina-Espeja, P., P.G. de Santos, and M. Alcalde, *Directed Evolution of Unspecific Peroxygenase*, in *Directed Enzyme Evolution: Advances and Applications*, M. Alcalde, Editor. 2017, Springer International Publishing: Cham. p. 127-143.

60. Martínez, A.T., et al., *Oxidoreductases on their way to industrial biotransformations.* Biotechnology Advances, 2017. **35**(6): p. 815-831.

61. Karich, A., et al., *Fungal Unspecific Peroxygenases Oxidize the Majority of Organic EPA Priority Pollutants.* Frontiers in Microbiology, 2017. **8**: p. 1463.

62. Hofrichter, M., et al., *Fungal Unspecific Peroxygenases: Heme-Thiolate Proteins That Combine Peroxidase and Cytochrome P450 Properties*, in *Monooxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome P450*, E.G. Hrycay and S.M. Bandiera, Editors. 2015, Springer International Publishing: Cham. p. 341-368.

63. Piontek, K., et al., *Structural basis of substrate conversion in a new aromatic peroxygenase: cytochrome P450 functionality with benefits.* J Biol Chem, 2013. **288**(48): p. 34767-76.

64. Molina-Espeja, P., et al., *Directed Evolution of Unspecific Peroxygenase from Agrocybe aegerita.* Applied and Environmental Microbiology, 2014. **80**(11): p. 3496-3507.

65. Gherib, R., H.M. Dokainish, and J.W. Gauld, *Multi-Scale Computational Enzymology: Enhancing Our Understanding of Enzymatic Catalysis.* International Journal of Molecular Sciences, 2014. **15**(1): p. 401-422.

66. Lodola, A. and A.J. Mulholland, *Computational Enzymology*, in *Biomolecular Simulations: Methods and Protocols*, L. Monticelli and E. Salonen, Editors. 2013, Humana Press: Totowa, NJ. p. 67-89.

67. Lonsdale, R., K.E. Ranaghan, and A.J. Mulholland, *Computational enzymology.* Chemical Communications, 2010. **46**(14): p. 2354-2372.

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

69. Shaw, D.E., et al., *Anton, a special-purpose machine for molecular dynamics simulation.* Commun. ACM, 2008. **51**(7): p. 91-97.

70. Wilson, S., *Many-body perturbation theory and its application to the molecular structure problem*, in *Chemical Modelling: Applications and Theory Volume 5*. 2008, The Royal Society of Chemistry. p. 208-248.

71. Acebes, S., et al., *Rational Enzyme Engineering Through Biophysical and Biochemical Modeling.* ACS Catalysis, 2016. **6**(3): p. 1624-1629.

72. Karabencheva, T. and C. Christov, *Chapter 1 - Structural and Computational Enzymology: Bringing Experiments and Computations Together*, in *Advances in Protein Chemistry and Structural Biology*, C. Christov and T. Karabencheva-Christova, Editors. 2012, Academic Press. p. 1-4.

73. Götz, A.W., M.A. Clark, and R.C. Walker, *An extensible interface for QM/MM molecular dynamics simulations with AMBER.* Journal of Computational Chemistry, 2014. **35**(2): p. 95-108.

74. Romero-Rivera, A., M. Garcia-Borras, and S. Osuna, *Computational tools for the evaluation of laboratory-engineered biocatalysts.* Chemical Communications, 2017. **53**(2): p. 284-297.

75. Lewars, E.G., *Introduction to Quantum Mechanics in Computational Chemistry*, in *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, E.G. Lewars, Editor. 2016, Springer International Publishing: Cham. p. 101-191.

76. Jensen, F., *Introduction to Computational Chemistry*. 2006: John Wiley \\& Sons.

77. Atkins, P.W. and R. Friedman, *Molecular quantum mechanics*. 5th ed. ed. 2011, Oxford: Oxford University Press.

78. Hehre, W.J., *A guide to molecular mechanics and quantum chemical calculations*. 2003: Irvine, CA : Wavefunction, [2003] ©2003.

79. Schlick, T., *Theoretical and Computational Approaches to Biomolecular Structure*, in *Molecular Modeling and Simulation: An Interdisciplinary*

Guide: An Interdisciplinary Guide, T. Schlick, Editor. 2010, Springer New York: New York, NY. p. 237-264.

80. Cramer, C.J., *Essentials of Computational Chemistry: Theories and Models (2nd Edition)*. 2004: John Wiley & Sons.

81. Young, D.C., *Ab Initio Methods*, in *Computational Chemistry*. 2001, John Wiley & Sons, Inc. p. 19-31.

82. Koch, W. and M.C. Holthausen, *The Hohenberg-Kohn Theorems*, in *A Chemist's Guide to Density Functional Theory*. 2001, Wiley-VCH Verlag GmbH. p. 33-40.

83. Young, D.C., *Density Functional Theory*, in *Computational Chemistry*. 2001, John Wiley & Sons, Inc. p. 42-48.

84. Koch, W. and M.C. Holthausen, *The Kohn-Sham Approach*, in *A Chemist's Guide to Density Functional Theory*. 2001, Wiley-VCH Verlag GmbH. p. 41-64.

85. Zhao, Y. and D.G. Truhlar, *A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions.* The Journal of Chemical Physics, 2006. **125**(19): p. 194101.

86. Young, D.C., *Semiempirical Methods*, in *Computational Chemistry*. 2001, John Wiley & Sons, Inc. p. 32-41.

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

88. Hug, S., *Classical Molecular Dynamics in a Nutshell*, in *Biomolecular Simulations: Methods and Protocols*, L. Monticelli and E. Salonen, Editors. 2013, Humana Press: Totowa, NJ. p. 127-152.

89. Lewars, E.G., *Molecular Mechanics*, in *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, E.G. Lewars, Editor. 2016, Springer International Publishing: Cham. p. 51-99.

90. Leach, A., *Molecular modelling : principles and applications*. 2009: Pearson Prentice Hall.

91. Saxena, A., et al., *Chapter 12 - The Basic Concepts of Molecular Modeling*, in *Methods in Enzymology*. 2009, Academic Press. p. 307-334.

92. Durrant, J.D. and J.A. McCammon, *Molecular dynamics simulations and drug discovery.* BMC Biology, 2011. **9**(1): p. 71.

93.   Zhang, J., et al., *Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents.* Journal of Chemical Theory and Computation, 2017. **13**(3): p. 1034-1043.

94.   Baker, C.M., *Polarizable force fields for molecular dynamics simulations of biomolecules.* Wiley Interdisciplinary Reviews: Computational Molecular Science, 2015. **5**(2): p. 241-254.

95.   Hospital, A., et al., *Molecular dynamics simulations: advances and applications.* Advances and Applications in Bioinformatics and Chemistry : AABC, 2015. **8**: p. 37-47.

96.   Hinchliffe, A., *Molecular modelling for beginners*. 2nd ed. ed. 2008, Oxford: Wiley-Blackwell.

97.   Earl, D.J. and M.W. Deem, *Monte Carlo Simulations*, in *Molecular Modeling of Proteins*, A. Kukol, Editor. 2008, Humana Press: Totowa, NJ. p. 25-36.

98.   Borrelli, K.W., et al., *PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique.* Journal of Chemical Theory and Computation, 2005. **1**(6): p. 1304-1311.

99.   Madadkar-Sobhani, A. and V. Guallar, *PELE web server: atomistic study of biomolecular systems at your fingertips.* Nucleic Acids Research, 2013. **41**(W1): p. W322-W328.

100.   Lucas, M.F. and V. Guallar, *Single vs. multiple ligand pathways in globins: A computational view.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2013. **1834**(9): p. 1739-1743.

101.   Hosseini, A., et al., *Atomic Picture of Ligand Migration in Toluene 4-Monooxygenase.* The Journal of Physical Chemistry B, 2015. **119**(3): p. 671-678.

102.   Borrelli, K.W., B. Cossins, and V. Guallar, *Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility.* Journal of Computational Chemistry, 2010. **31**(6): p. 1224-1235.

103.   Cossins, B.P., A. Hosseini, and V. Guallar, *Exploration of Protein Conformational Change with PELE and Meta-Dynamics.* Journal of Chemical Theory and Computation, 2012. **8**(3): p. 959-965.

104.   Cabeza de Vaca, I., M.F. Lucas, and V. Guallar, *New Monte Carlo Based Technique To Study DNA–Ligand Interactions.* Journal of Chemical Theory and Computation, 2015. **11**(12): p. 5598-5605.

105.   Takahashi, R., V.A. Gil, and V. Guallar, *Monte Carlo Free Ligand Diffusion with Markov State Model Analysis and Absolute Binding Free Energy Calculations.* Journal of Chemical Theory and Computation, 2014. **10**(1): p. 282-288.

106. Atilgan, A.R., et al., *Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model.* Biophysical Journal. **80**(1): p. 505-515.

107. Kaminski, G.A., et al., *Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides.* The Journal of Physical Chemistry B, 2001. **105**(28): p. 6474-6487.

108. Bashford, D. and D.A. Case, *Generalized born models of macromolecular solvation effects.* Annu Rev Phys Chem, 2000. **51**: p. 129-52.

109. Gu, J. and P.E. Bourne, *Structural Bioinformatics.* 2011: John Wiley & Sons. 1064.

110. Meng, X.-Y., et al., *Molecular Docking: A powerful approach for structure-based drug discovery.* Current computer-aided drug design, 2011. **7**(2): p. 146-157.

111. Warshel, A. and M. Levitt, *Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme.* Journal of Molecular Biology, 1976. **103**(2): p. 227-249.

112. Groenhof, G., *Introduction to QM/MM Simulations*, in *Biomolecular Simulations: Methods and Protocols*, L. Monticelli and E. Salonen, Editors. 2013, Humana Press: Totowa, NJ. p. 43-66.

113. van der Kamp, M.W. and A.J. Mulholland, *Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology.* Biochemistry, 2013. **52**(16): p. 2708-2728.

114. Senn, H.M. and W. Thiel, *QM/MM Methods for Biomolecular Systems.* Angewandte Chemie International Edition, 2009. **48**(7): p. 1198-1229.

115. Ryde, U., *Chapter Six - QM/MM Calculations on Proteins*, in *Methods in Enzymology*, A.V. Gregory, Editor. 2016, Academic Press. p. 119-158.

116. *QSite*. 2017, Schrödinger, LLC: New York, NY.

117. Berg, J.M., J.L. Tymoczko, and L. Stryer, *Biochemistry.* 2011: Macmillan Higher Education. 1120.

118. Ichiye, T., *Electron Transfer Proteins: Overview*, in *Encyclopedia of Biophysics*, G.C.K. Roberts, Editor. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 614-621.

119. Wallrapp, F., D. Masone, and V. Guallar, *Electron Transfer in the P450cam/PDX Complex. The QM/MM e-Pathway.* The Journal of Physical Chemistry A, 2008. **112**(50): p. 12989-12994.

120. Guallar, V. and F. Wallrapp, *Mapping protein electron transfer pathways with QM/MM methods.* Journal of The Royal Society Interface, 2008. **5**(Suppl 3): p. 233.

121. Marcus, R.A. and N. Sutin, *Electron transfers in chemistry and biology.* Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics, 1985. **811**(3): p. 265-322.

122. Cabeza de Vaca, I., S. Acebes, and V. Guallar, *Ecoupling server: A tool to compute and analyze electronic couplings.* Journal of Computational Chemistry, 2016. **37**(18): p. 1740-1745.

123. Saen-Oon, S., M.F. Lucas, and V. Guallar, *Electron transfer in proteins: theory, applications and future perspectives.* Physical Chemistry Chemical Physics, 2013. **15**(37): p. 15271-15285.

124. Chakraborty, T., P. Ranjan, and A. Pandey, *Computational Chemistry Methodology in Structural Biology and Materials Sciences.* 2017: CRC Press. 370.

125. Richter, F., et al., *De Novo Enzyme Design Using Rosetta3.* PLOS ONE, 2011. **6**(5): p. e19230.

126. Bolon, D.N. and S.L. Mayo, *Enzyme-like proteins by computational design.* Proceedings of the National Academy of Sciences, 2001. **98**(25): p. 14274-14279.

127. Kries, H., R. Blomberg, and D. Hilvert, *De novo enzymes by computational design.* Current Opinion in Chemical Biology, 2013. **17**(2): p. 221-228.

128. Huang, P.-S., S.E. Boyken, and D. Baker, *The coming of age of de novo protein design.* Nature, 2016. **537**: p. 320.

129. Faiella, M., et al., *An artificial di-iron oxo-protein with phenol oxidase activity.* Nature Chemical Biology, 2009. **5**: p. 882.

130. Mak, W.S. and J.B. Siegel, *Computational enzyme design: Transitioning from catalytic proteins to enzymes.* Current Opinion in Structural Biology, 2014. **27**(Supplement C): p. 87-94.

131. Bornscheuer, U.T. and M. Pohl, *Improved biocatalysts by directed evolution and rational protein design.* Current Opinion in Chemical Biology, 2001. **5**(2): p. 137-143.

132. Yang, W. and L. Lai, *Computational design of ligand-binding proteins.* Current Opinion in Structural Biology, 2017. **45**(Supplement C): p. 67-73.

133. Leaver-Fay, A., et al., *Chapter nineteen - Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules*, in

Methods in Enzymology, M.L. Johnson and L. Brand, Editors. 2011, Academic Press. p. 545-574.

134. Gainza, P., et al., *OSPREY: Protein Design with Ensembles, Flexibility, and Provable Algorithms.* Methods in enzymology, 2013. **523**: p. 87-107.

135. Moretti, R., et al., *Rosetta and the Design of Ligand Binding Sites*, in *Computational Design of Ligand Binding Proteins*, B.L. Stoddard, Editor. 2016, Springer New York: New York, NY. p. 47-62.

136. Kiss, G., et al., *Computational Enzyme Design.* Angewandte Chemie International Edition, 2013. **52**(22): p. 5700-5725.

137. Kiss, G., et al., *Evaluation and ranking of enzyme designs.* Protein Science : A Publication of the Protein Society, 2010. **19**(9): p. 1760-1773.

138. Jiménez-Osés, G., et al., *The role of distant mutations and allosteric regulation on LovD active site dynamics.* Nature Chemical Biology, 2014. **10**: p. 431.

139. Chovancova, E., et al., *CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures.* PLOS Computational Biology, 2012. **8**(10): p. e1002708.

140. Sehnal, D., et al., *MOLE 2.0: advanced approach for analysis of biomacromolecular channels.* Journal of Cheminformatics, 2013. **5**: p. 39-39.

141. Le Guilloux, V., P. Schmidtke, and P. Tuffery, *Fpocket: An open source platform for ligand pocket detection.* BMC Bioinformatics, 2009. **10**(1): p. 168.

142. Brady, G.P. and P.F.W. Stouten, *Fast prediction and visualization of protein binding pockets with PASS.* Journal of Computer-Aided Molecular Design, 2000. **14**(4): p. 383-401.

143. Grisewood, M.J., et al., *OptZyme: Computational Enzyme Redesign Using Transition State Analogues.* PLOS ONE, 2013. **8**(10): p. e75358.

144. Khersonsky, O., et al., *Optimization of the in silico designed Kemp eliminase KE70 by computational design and directed evolution.* Journal of molecular biology, 2011. **407**(3): p. 391-412.

145. Zheng, F., et al., *Most Efficient Cocaine Hydrolase Designed by Virtual Screening of Transition States.* Journal of the American Chemical Society, 2008. **130**(36): p. 12148-12155.

146. Monza, E., et al., *Insights into Laccase Engineering from Molecular Simulations: Toward a Binding-Focused Strategy.* The Journal of Physical Chemistry Letters, 2015. **6**(8): p. 1447-1453.

147. Lutz, S., *Beyond directed evolution—semi-rational protein engineering and design.* Current Opinion in Biotechnology, 2010. **21**(6): p. 734-743.

148. Davids, T., et al., *Strategies for the discovery and engineering of enzymes for biocatalysis.* Current Opinion in Chemical Biology, 2013. **17**(2): p. 215-220.

149. Carvalho, A.T.P., et al., *Challenges in computational studies of enzyme structure, function and dynamics.* Journal of Molecular Graphics and Modelling, 2014. **54**(Supplement C): p. 62-79.

150. Damborsky, J. and J. Brezovsky, *Computational tools for designing and engineering enzymes.* Current Opinion in Chemical Biology, 2014. **19**: p. 8-16.

151. Field, M.J., *A Practical Introduction to the Simulation of Molecular Systems*. 2 ed. 2007, Cambridge: Cambridge University Press.

152. Trewin, A., *Molecular Modelling for Beginners.* Chromatographia, 2010. **71**(1): p. 175-175.

153. Gerlt, J.A. and P.C. Babbitt, *Enzyme (re)design: lessons from natural evolution and computation.* Current Opinion in Chemical Biology, 2009. **13**(1): p. 10-18.

154. Beratan, D.N., et al., *Electron-tunneling pathways in proteins.* Science, 1992. **258**(5089): p. 1740.

155. Johnston, J.B., et al., *Structural control of cytochrome P450-catalyzed ω-hydroxylation.* Archives of Biochemistry and Biophysics, 2011. **507**(1): p. 86-94.