



UNIVERSITAT DE
BARCELONA

Computational design of oxidoreductases for industrial applications

Ferran Sancho Jodar

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Computational design of oxidoreductases for industrial applications

Ferran Sancho Jodar



UNIVERSITAT DE
BARCELONA

Programa de Doctorat en Química Teòrica i Modelització Computacional

Departament de Química Física

Facultat de Química

Universitat de Barcelona

Computational design of oxidoreductases for industrial applications

Memòria presentada per:

FERRAN SANCHO JODAR

Barcelona, setembre del 2019

per a optar al grau de Doctor per la Universitat de Barcelona

Treball realitzat a Barcelona Supercomputing Center – Centre Nacional de Supercomputació (BSC-CNS).

Dirigida per:

Víctor Guallar Tasies

Maria Fatima Lucas

Barcelona Supercomputing Center

Zymvol Biomodeling SL

Tutoritzada per:

Jaime Rubio Martínez

Universitat de Barcelona

CONTENTS

ACKNOWLEDGEMENTS	i
LIST OF PUBLICATIONS	ii
OUTLINE OF THE THESIS	iii
ABBREVIATIONS	iv
I. INTRODUCTION	1
1. Preface	3
1.1. The motivation behind this thesis	3
1.1.1. Enzymes in nature	3
1.1.2. Use of enzymes for industrial purposes	4
1.2. Introduction to enzyme engineering	4
1.2.1. Experimental strategies.....	4
1.2.2. Computational strategies	7
1.3. Applications of enzymes	8
1.4. The future of enzyme engineering	8
2. Systems of study	10
2.1. General structure of proteins	10
2.2. Flavoproteins	11
2.2.1. Aryl-alcohol oxidase.....	15
2.2.2. Hydroxymethylfurfural Oxidase	17
2.3. Laccases	17
2.3.1. The use of a photosensitizer with laccases	20
3. Theory	22
3.1. Protein related theory	22
3.1.1. Enzyme kinetics.....	22
3.1.2. Electron transfer in proteins	24
3.2. Computational enzyme modelling	25
3.2.1. Quantum Mechanics	25
3.2.2. Molecular Mechanics	29

3.2.3. Quantum Mechanics / Molecular Mechanics	32
3.2.4. Geometry optimization	33
3.2.5. Molecular dynamics	34
3.2.6. Monte Carlo	36
3.2.7. Docking.....	37
3.2.8. Protein Energy Landscape Exploration algorithm.....	37
II. OBJECTIVES	41
III. RESULTS	45
1. Flavoproteins	47
1.1. Article I	47
1.2. Article II	69
1.3. Article III	87
1.4. Article IV	109
2. Laccases	129
2.1. Article V	129
IV. DISCUSSION.....	149
1. Results, summary and discussion.....	151
1.1. Secondary alcohol oxidation by means of flavoproteins	152
1.1.1. F397 variants for ligand diffusion	154
1.1.2. Combinatorial mutagenesis for residues 500 and 501 in AAO.....	157
1.1.3. Rationalization of the mutations introduced in FX9 variant of AAO.....	159
1.1.4. Tuning FAD potential by the evaluation of residue electrostatics.....	162
1.2. Improvement of full HMF oxidation.....	164
1.3. Laccases	170
1.3.1. Interaction of laccases with a nanotube.....	174
V. CONCLUSIONS.....	177
VI. REFERENCES.....	181

ACKNOWLEDGEMENTS

En primer lloc, m'agradaria donar les gràcies tant a en Víctor Guallar com a la Fátima Lucas per haver-me donat l'oportunitat de fer el doctorat al BSC. Vull fer una menció especial a en Jaime Rubio, qui em va introduir en el món dels enzims i em va passar l'oferta de treball. També agrair a INDOX i Anaxomics pel projecte i el suport financer.

Aquests anys he tingut el plaer de compartir un espai de treball amb un grup i un ambient immillorable, d'on no només m'emporto bons records sinó molt bons amics. En aquests últims dies ja no queda cap dels membres del grup en el que vaig arribar, però tot i així la dinàmica no ha canviat gens. Gràcies a tots els que heu compartit amb mi aquests anys al BSC.

També vull donar les gràcies a la Joelle Pelletier, per fer possible la meva estada a Montréal, juntament amb PROTEO, FRQNT i tota la gent que vaig conèixer allà.

Arribar fins aquí no ha estat fàcil i no sé si hagués estat possible sense l'ajuda de l'Hèctor i en Gerard. Sempre us agrairé tot el que m'heu ajudat durant la carrera i el màster.

I, per descomptat, gràcies a la meva família: als meus avis, la meva mare i la meva germana, per acompanyar-me al llarg de tot aquest camí.

M'agradaria fer especial menció a la Fátima i a l'Emanuele, gràcies per tot el que m'heu ensenyat, tant professionalment com personalment.

Durant la tesi hi ha hagut moments complicats, però hi ha gent que ha estat al meu costat, que s'ha preocupat per mi i m'ha ajudat en tot el possible i més. De tot cor, moltíssimes gràcies a tots vosaltres.

I en resum, gràcies a tota la gent que ha passat per la meva vida durant aquests anys, perquè d'una manera o altra formeu part d'aquesta tesi.

Gràcies a tots.

LIST OF PUBLICATIONS

This thesis consists of a compendium of the following articles, included in this order:

- I. Carro J, Amengual-Rigo P, **Sancho F**, Medina M, Guallar V, Ferreira P, Martínez AT. Multiple implications of an active site phenylalanine in the catalysis of aryl-alcohol oxidase. *Sci Rep.* 2018;8: 8121.
- II. **Sancho F**, Serrano A, Viña-González J, Carro J, Alcalde M, Guallar V, Martínez AT. Switching the substrate preference of fungal aryl-alcohol oxidase: towards stereoselective oxidation of secondary benzyl alcohols. *Catalysis Science & Technology.* 2019. pp. 833–841. doi:10.1039/c8cy02447
- III. Viña-Gonzalez J, Jimenez-Lalana D, **Sancho F**, Serrano A, Martínez AT, Guallar V, Alcalde M. Structure-Guided Evolution of Aryl Alcohol Oxidase from *Pleurotus eryngii* for the Selective Oxidation of Secondary Benzyl Alcohols. *Advanced Synthesis & Catalysis.* 2019. doi:10.1002/adsc.201900134
- IV. **Sancho F**, Cañellas M, Guallar V and Lucas F. (2018) Towards the efficient production of FDCA: Molecular Modelling and In Silico Design of AaeUPO, HMFO and AAO enzymes. (*Draft manuscript*)
- V. Robert V, Monza E, Tarrago L, **Sancho F**, De Falco A, Schneider L, Ngoutane, EN, Mekmouche Y, Pailley PR, Jalila Simaan, A; Guallar V, Tron T. Probing the Surface of a Laccase for Clues towards the Design of Chemo-Enzymatic Catalysts. *ChemPlusChem.* 2017. pp. 607–614. doi:10.1002/cplu.201700030

Moreover, the following publications have been developed during this period, but not included in this thesis:

- VI. **Sancho F.**, Guallar V, Tron T, *et al.* Direct wiring of azide-modified laccases on carbon nanotubes (*Draft manuscript*)
- VII. Alejaldre L, **Sancho F**, Guallar V, Pelletier J, *et al.* Robustness and protein epistasis in TEM-1 β -lactamase (*Draft manuscript*)

OUTLINE OF THE THESIS

This thesis has been assembled as a compendium of publications, submitted manuscripts and draft manuscripts. It is structured in 6 chapters: It commences with a general introduction (chapter I) followed by the objectives of the thesis (chapter II). Results are presented as articles from Article I to V (chapter III) that will be later summarized and discussed in chapter IV. Finally, this thesis is brought to a close with the conclusions (chapter V) and the references included in the text (chapter VI).

Herein, I cover the study of enzymes for industrial applications and, in particular, somehow involved in the lignin degradation process. For this reason, the introduction (chapter I) sets up a prologue of the importance of enzymes in nature and in the industry, exposing the applications of two particular families of enzymes: flavoproteins and laccases. Emphasis will also be put on the different methods used to study these systems.

The articles here incorporated have been classified according to the enzymatic system studied and thus, their applications in industry: Articles I-IV for flavoproteins and article V for laccases. Article VI (not included as a manuscript) is a continuation of article V, but my personal contribution is only reflected in the supporting information. For this reason, I decided not to include it, but comment on the results as a sub-section in the discussion (chapter IV) instead. Finally, article VII was not included, as it is off the topic.

Last but not least, it is important to point out that all the work here presented has been the result of the collaboration of several scientists, from diverse research groups and expertise areas. Even though articles have been included as a whole in chapter III, I have only contributed to the computational section of these publications, and this is the part that will be discussed in chapter IV.

ABBREVIATIONS

AAO	Aryl-alcohol oxidase
ANI	<i>p</i> -anisyl
ANM	Anisotropic Network Model
CV	Collective variable
DFF	2,5-diformylfuran
DFT	Density Functional Theory
FDCA	2,5-furandicarboxylic acid
FES	Free Energy Surface
FFCA	2,5-formylfurancarboxylic acid
HMF	5-hydroxymethylfurfural
HMFO	5-hydroxymethylfurfural Oxidase
INDOX	Industrial Oxidoreductases
k_{cat}	Catalytic constant
k_{cat}/k_M	Catalytic efficiency
K_M	Michaelis-Menten constant
LCAO	Linear Combination of Atomic Orbitals
MC	Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics
PELE	Protein Energy Landscape Exploration
PES	Potential Energy Surface
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
RMP	(R)-1-(<i>p</i> -methoxyphenyl) ethanol
SMP	(S)-1-(<i>p</i> -methoxyphenyl) ethanol
TS	Transition State
UPO	Unspecific Peroxygenase

I. INTRODUCTION

1. Preface

1.1. The motivation behind this thesis

1.1.1. Enzymes in nature

Nature is extraordinary. The concept of life understood as a complex system of biochemical reactions is already fascinating. But in a neutral aqueous solution at room temperature, some of these reactions are so slow that they would not occur under conditions compatible with life (Radzicka and Wolfenden 1995). The main difference lies in the presence (or absence) of enzymes, biocatalysts that speed up the rates of these biochemical reactions. It has been extrapolated that the half-time of some decarboxylation reactions would be of millions (Miller and Wolfenden 2002) or even billions (Lewis and Wolfenden 2008) of years. And, remarkably, enzymes are able to reduce these half-time values to milliseconds. To evaluate the different efficiencies of these reactions (with and without enzymes), Wolfenden et al. used the expression “catalytic proficiency”, reaching values up to $2.5 \times 10^{24} \text{ M}^{-1}$, considered the largest rate enhancement of a protein without cofactors. Consequently, the role of enzymes is of great importance.

Intuitively, one would think about enzymes as perfectly ensembled machines which, optimized throughout evolution, have reached maximal performance. This can only be accomplished when their kinetic efficiency values (k_{cat}/K_M) reach the physical limit of diffusion rate, what would be considered a perfect enzyme (Goldsmith and Tawfik 2017). However, 98% of all enzymes are at least 10 times slower than the perfect enzyme (Bar-Even et al. 2015) and the average kinetic efficiencies in the literature are 4 orders of magnitude lower (Bar-Even et al. 2011). What it is not being considered is that these enzymes were optimized in their cellular context, where their catalytic efficiency values are limited by flux-balance factors, to avoid toxic levels of some products, for example (Reznik, Mehta, and Segrè 2013; Sajitz-Hermstein and Nikoloski 2016). As a result, enzymes are very selective about the substrate they accept, and even promiscuous enzymes have different efficiencies towards different substrates (Bar-Even et al. 2011). In order to be regulated by flux-balance factors, enzymes also need to be turned on and off, so they can be inhibited by their own products. Additionally, they often stand against environments different from the ones they have evolved for. In other words, enzymes in living organisms have been optimized for their needs, not to maximize their productivity.

1.1.2. Use of enzymes for industrial purposes

Contrarily, industries have different goals: seeking for maximal catalytic efficiency peaks or improved selectivity or stability are just some of the main industrial interests. Therefore, in order to accomplish these objectives, *the end justifies the means*, which symbolizes the variety of strategies to increase enzymatic performance. While disregarding “*in vivo*” limitations, other obstacles arise during the scale-up feasibility, like long-term stability under process conditions, difficulties in recovery and recycling or the production of enzyme in sufficient quantity to make it economically viable. Besides adjusting the working environmental conditions that affect the activity of an enzyme (temperature, pH, concentration, ionic strength...), the procedures of improving enzyme performances can be very different, according to the difficulty that is being faced. For instance, enzyme immobilization can tackle destabilization or short lifespan problems (Mohamad et al. 2015), or even increase the activity (Mateo et al. 2007). Other strategies to enhance activity or stability include the introduction or engineering of chaperones (Mack and Shorter 2016; Saio et al. 2018), proteins that assist, for better or worse, folding or assembly, or covalent tethering of small molecules that work as activators to an enzyme (Darby et al. 2017). Overall, protein engineering has settled as a method to overcome industrial enzymes' limitations by modifying the protein sequence to obtain the desired result. This uprising interest in enzyme engineering for biocatalysts is becoming the fastest growing in the biotechnology sector (Barrozo et al. 2012). Some industrial enzyme applications have been collected in Table 1.

1.2. Introduction to enzyme engineering

Since the first time that a protein was engineered, many efforts have been dedicated in order to tailor enzymes a la carte for diverse applications. Although this thesis is purely based on computational work, there is a strong synergy between experiments and simulations in the articles here presented. For this reason, a brief introduction of the evolution of both experimental and computational strategies are presented.

1.2.1. Experimental strategies

The experiments with recombinant DNA carried out during the early 1970s by Stanley Cohen, Herbert Boyer and their team laid the first stone of enzyme engineering, allowing

Table 1. Major industrial enzyme applications (from Jemli et al. 2016; Li et al. 2012).

	Industry	Enzyme	Applications
Technical industry	Laundry detergents	Amylases Lipases Proteases Xylanases	Removal of resistant starch residues Lipid stain removal Protein-based stain removal Plant-based stain removal
	Textile industry	Amylases Cellulases Laccases Peroxidases	Starch removal from woven fabrics Denim bleaching, textile softener Decolorization and detoxification of effluents from textile Dye excess removal
	Pulp and paper industry	Cellulases Flavoproteines Laccases Lipases	Making fibres flexible Involved in lignin degradation Improving brightness and removal of lignin from fibres Control pitch in pulping process
	Biorefinery	Cellulases, hemicellulases Lipases	Degrading efficiently the lignocellulotic materials for ethanol production Biodiesel production by transesterification
Food industry	Dairy industry	β -galactosidases Lipases Proteases	Avoiding lactose intolerance Cheese flavour Milk clotting, flavor
	Starch industry	α -amylases, amyloglucosidases Cyclodextrin glycosyltransferases Glucose isomerases	Conversion of starch to glucose syrup Cyclodextrins production and carbohydrate glycosylation Production of high fructose syrup
	Baking industry	α -amylases Glucose oxidases Lipases Lipoxygenases Proteases	Controlling the volume and crumb structure of bread Improving stability of gas cells in dough Dough strengthening Dough strengthening, bread whitening Biscuits production
	Juice industry	Amylases Laccases Naringinases, limoninases Pectinases	Clarifying cloudy juice Phenol derivative removal Decreasing bitterness in citrus juices Increasing juice production yield
	Brewing industry	β -glucanases Pentosanases, xylanases Proteases Tannases	Decreasing viscosity and improving wort separation Improving extraction and beer filtration Improving yeast growth Clarifying agent
	Fat and oil industry	Lipases Phospholipases	Flavour production Lyso-lecithin production
	Functional food industry	Pectinases Phytases Rhamnosidases	Obtaining solubilized dietary fiber via treatment of potato pulp Dephosphorylation of phytic acid to enhance iron absorption Enhancing bioavailability of the citrus flavonoid hesperetin
Animal feed industry	β -glucanases Phytases Tannases Xylanases	Increasing animal feed digestibility Releasing phosphorus and increasing availability of cations Hydrolysis of tannins and gallic acid esters Degrading fiber in viscous diets and increasing nutritive value	
Organic synthesis industry	Laccases Lipases Monooxygenases Transaminases	Oxidation of phenol derivatives for synthetic applications Resolution of chiral compounds Synthesis of optically pure epoxides Resolution of racemic amines and direct chiral synthesis	
Cosmetics industry	Glucose oxidases Oxidases, peroxidases Papain, bromelain, subtilisin	Toothpastes and mouthwashes Hair dyeing Giving peeling effects in skin care	

the mass production of proteins (Cohen et al. 1973). A few years later, Genentech applied for the first time recombinant DNA techniques to produce synthetic human insulin for industrial applications (Goeddel et al. 1979). Together with an extensive list of advances achieved in molecular biology during these years, such as DNA amplification by polymerase chain reaction or DNA site-specific sequence manipulation with endonucleases and ligases, enabled the precise replacement of specific amino acid in the protein sequence (Winter et al. 1982; Wilkinson et al. 1983). This attracted all the attention towards the role of amino acids as individuals on the protein function, structure or stability. And soon, in 1985, site-directed mutagenesis was used for the first time as a tool to overcome a commercial limitation of an enzyme (Estell, Graycar, and Wells 1985). Since this technique is rational-design based, it relies on the previous knowledge of the enzyme and its reaction mechanism, in order to know which amino acid to target, which is an important limitation.

Already in the late 80s, the first protocols that mimic natural evolution appeared under the name of directed evolution, based on random mutagenesis (Cadwell and Joyce 1992). Here, the best variants were selected for future rounds of evolution, accumulating beneficial mutations each round. However, this protocol has its own disadvantages, since beneficial amino acid substitutions must be found independently, meaning that variants with synergistic effects (epistasis) could be invisible during the selection after each round. Nevertheless, directed evolution has been improved over the last two decades, and nowadays it remains a key method in the field of enzyme engineering. So much so that in 2018 Frances Arnold was awarded half of the Nobel Prize in Chemistry for her work in directed evolution (Arnold et al. 1993).

The lack of previous knowledge in random mutagenesis experiments could lead to the requirement of enormous combinatorial libraries before finding a beneficial mutation (or set of mutations), and random mutagenesis of each residue of a protein can be highly expensive and time-consuming. So, part of the success of emerging directed evolution strategies is due to the use of smaller libraries taking advantage of information from the protein. These libraries contain mutations in the sequence of preselected promising positions, obtained by rational design, limiting the variability but enriching (in theory) the ratio of positive results. The combination of rational design and random mutagenesis is referred as semi-rational design (Chica, Doucet, and Pelletier 2005). And here, the potential of computational tools is noteworthy.

1.2.2. Computational strategies

While bioinformatics is still classified as an emergent field, computers have been present in molecular biology since the early 1960s (Hagen 2000). And already in 1985, DeGrado and coworkers applied for the first time computational methods to model and design a 17-residue helical peptide, a calmodulin inhibitor (DeGrado et al. 1985). It included side-chain positioning, minimization and the protein-protein surface interaction based on electrostatic potential surfaces and structural modeling (O'Neil and DeGrado 1985). Since then, computational methods have been a very useful modeling tool in the field, for instance, in the determination of the reaction mechanism of a protein (Caldararu et al. 2018; Berraud-Pache, Garcia-Iriepa, and Navizet 2018), protein structure predictions from homologous structures (Tong and Ranganathan 2013), based on the fact that structure is more conserved than the sequence (Bajaj 1984) or even protein structure predictions from the sequence, without any previous knowledge (AlQuraishi 2019). Since it is still challenging to study the protein-structure-function relation experimentally, computational methods play a key role here, with a considerable improvement in quality model predictions in the last years (Moult et al. 2018).

Modeling tools can further be exploited for enzyme redesign: while rational redesign for activity or stability predictions have been present (Steiner and Schwab 2012; Korendovych 2018), the exponential growth of entries in the Protein Data Bank, together with an increment of computer power of the past few decades, encouraged the development of new computational strategies to perform screening of mutations. These strategies were focused to either reduce the sequence space in experimental evolution studies or even as a direct *in silico* directed evolution (Allen, Nisthal, and Mayo 2010; Fox and Huisman 2008; Gustafsson, Govindarajan, and Minshull 2003; Voigt et al. 2001). The limits of computation did not stop there, and were soon used for *de novo* design: to design enzymes from scratch, that are not known to exist in nature (Röthlisberger et al. 2008; Jiang et al. 2008). As an example, David Baker's lab, considered a pioneer in the protein *de novo* design, designed a stereoselective catalysts for Diels-Alder reaction not naturally occurring in enzymes (Siegel et al. 2010). However, enhancements obtained are still modest compared to proficiencies of naturally occurring enzymes.

The explosion of interest and the huge number of advances in protein engineering have a very close relation with the remarkable commercial success of several enzyme-based industries all around the world, and the global enzyme market keeps growing every year (Chapman, Ismail, and Dinu 2018). The advantages of enzymes were soon seen to hold

potential in reducing operating costs for some industries and therefore they are investing a lot of effort in finding new applications by enzyme engineering.

1.3. Applications of enzymes

Even though naturally occurring enzymes have been used for food production and other commodities, enzyme industry has experienced a fast growth during the last decades. Enzymatic catalysis has been scaled up for pharmaceutical, food and beverage industries (Kapoor, Rafiq, and Sharma 2017; Huisman and Collier 2013), attracting the attention mainly from the energy sector, where further enhancements are still needed to make biocatalysts economically competitive (Fei et al. 2014; Asgher et al. 2014).

In this thesis, I have worked in the INDOX (INDustrial OXidoreductases <http://www.indoxproject.eu/>) project under the topic of “optimal and cost-effective industrial biocatalysts”. This project was mainly oriented towards the optimization of enzymes for industrial purposes and in this thesis, I have focused my investigations on oxidoreductases, to be applied in lignin and lignin-derived products. These oxidoreductases have an interest in pulp and paper industries and will be also of a huge interest for future lignocellulose bio-refineries (Bujanovic et al. 2012; Amidon et al. 2008).

In this thesis, I have explored different enzymes, particularly flavoproteins and laccases, by means of computational techniques for protein modelling and design. In the next chapter, details of the structure and mechanism of each system will be presented.

1.4. The future of enzyme engineering

As a personal opinion, the ultimate goal of enzyme engineering would be a personalized *de novo* design: the creation of a new enzyme specifically designed for each case. But, at this moment, we are far from this point.

The nearer future is expected to be conditioned by artificial intelligence algorithms. There is a huge amount of data from enzymes and sooner than later we will see new algorithms taking advantage of this resource. In fact, this was already seen in 2018 when Google's AI DeepMind was placed first in CASP protein structure prediction.

Also, it is predictable that the existing gap between computational enzyme engineering and experimentalist will be shrinking after the years. The availability of computer power (more resources and cheaper prices) is allowing the use of more computationally expensive algorithms, even to smaller research groups or companies. Thus, predictions will become more accurate, building trust bridges between worlds.

2. Systems of study

In this chapter, a summary of the structural information of the enzymes under study, needed to better understand the content of this thesis, is provided. Firstly, however, a very brief introduction to general protein structure is included. Then, the reaction mechanism, as well as the role of some important residues and some structural key points for each system, are introduced.

2.1. General structure of proteins

Proteins are large macromolecules formed through the combination of 20 natural amino acids (Figure 1). These amino acids, also called residues, are assembled as a chain. The link between amino acids is called the peptidic bond, which is the chemical bond formed between the carboxylic end of one residue and the amino group of the subsequent amino acid. These two groups, together with the alpha carbon comprised between them constitute the backbone of the protein. The remaining atoms are the side chains of the residue.

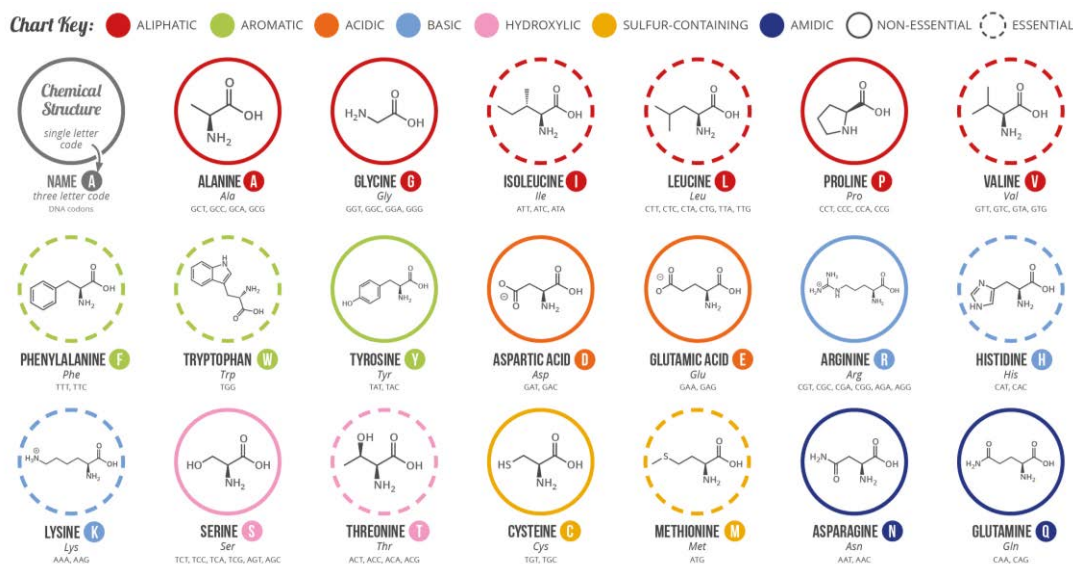


Figure 1. Chemical structure of the 20 natural amino acids with their three letter and single letter codes (modified from <https://www.compoundchem.com/2014/09/16/aminoacids/>).

Amino acids are usually classified according to the properties of their side chain, given by their composition and structure. Thus, we can find amino acids electrically charged, polar, apolar or hydrophobic and the special cases. This will determine the interaction of amino acids with other residues or substrates.

The structure of a protein can be subdivided into four levels (Figure 2). The primary structure refers to the early mentioned chain, the sequence of amino acids, where the function of the protein is encoded. The polypeptide backbone chain acquires local patterns with the form of helices, sheets or loops, conforming the secondary structure. Their distribution in the three-dimensional space, folding into a globular structure, for example, and the assembly between different proteins are the tertiary and quaternary structures respectively.

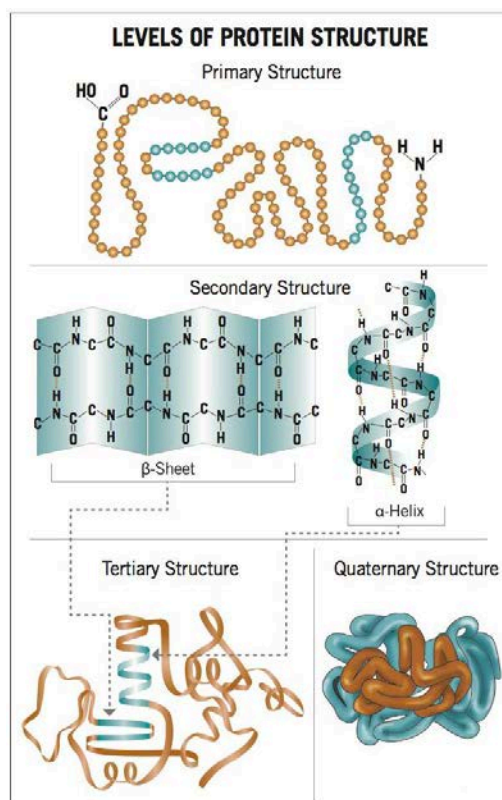


Figure 2. Primary, secondary, tertiary, and quaternary protein structures (modified from *Particle Sciences - Technical Brief: 2009: Volume 8*).

2.2. Flavoproteins

Flavoproteins is the name given to those proteins with a flavin adenine dinucleotide (FAD) or in lesser amounts, flavin mononucleotide (FMN) as a prosthetic group or cofactor. These molecules are nucleic acids derivatives of riboflavin, which is also known as vitamin B₂. Most of the flavoprotein crystal structures contain non-covalently bound FAD or FMN but strongly attached in the active site, usually buried inside the protein. However, while it is believed that the covalent bond could increase the oxidative

capabilities of the cofactor by increasing its redox potential (Fraaije et al. 1999), other functions have been attributed to it as well, such as structure stabilization or a contribution to the substrate binding (Dourado, Swart, and Carvalho 2018).

The flavin structure consists of very distinct parts (Figure 3) i) The isoalloxazine ring, which is the aromatic portion, common for all the flavins, and ii) the ribityl chain. The type of flavin depends on the substituent attached here, hydrogen for riboflavin, phosphate ion for FMN and adenosine diphosphate (ADP) for FAD.

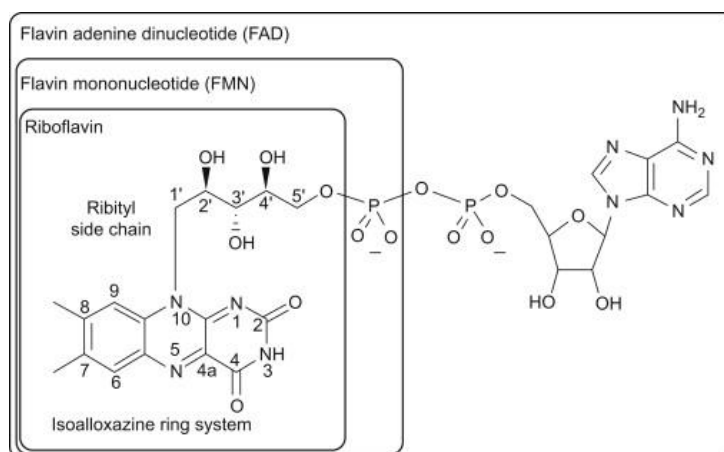


Figure 3. Different structural parts of flavins (from Mazzotta et al. 2014).

The ribityl part plays an important role in the molecular recognition and binding of the cofactor. On the contrary, the isoalloxazine ring participates directly in the reaction mechanism. It is precisely this aromatic system what makes this cofactor special: it has the capability to participate in reactions involving either one-electron or two-electron transfer processes, which makes them fundamental as mediators between processes with a different number of electrons involved in biological systems (Edwards 2014). Thus, flavin molecules can adopt three different redox states: oxidized, semiquinone (one-electron reduced) and two-electron reduced (Figure 4). For this reason, flavoproteins can catalyse a variety of reactions (Ghisla and Massey 1989). However, and from now on, flavoprotein reactivity will be focused on alcohol oxidations by FAD cofactors, since it is the reaction mechanism performed by the systems studied here.

As in any redox reaction, processes catalysed by flavoenzymes are split into a reductive half-reaction and oxidative half-reaction. In the former, the substrate is oxidized with the complementary reduction of the FAD. In the latter, the reduced flavin is oxidized by an electron acceptor and recovered to start the catalytic cycle again. Molecular oxygen acts as the final electron acceptor of the flavoproteins studied in this thesis (Figure 5).

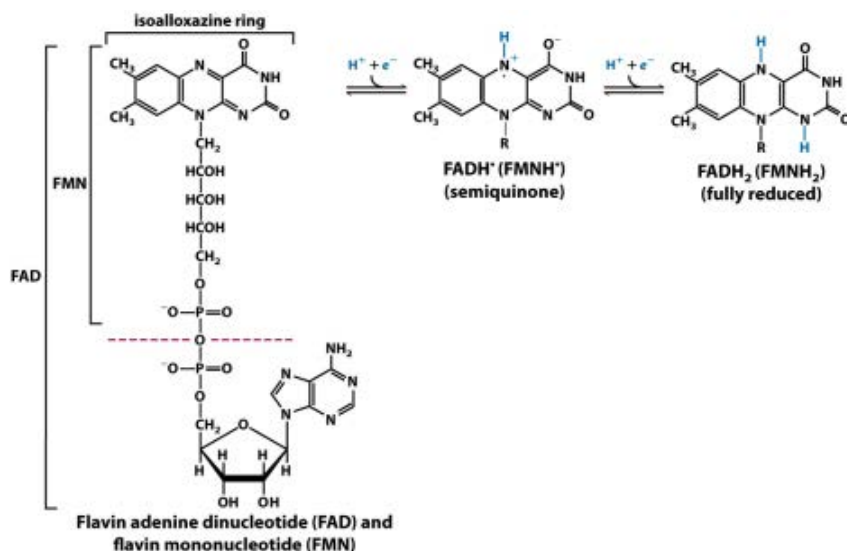


Figure 4. Different oxidation states that can be adopted by flavins (from Nelson et al. 2008).

The flavin cofactor is not self-sufficient to catalyse alcohol oxidations. It requires the presence of a residue that acts as a base or proton acceptor. In the flavoenzymes studied in this thesis, this role is fulfilled by a histidine adjacent to the isoalloxazine ring. In the reductive half-reaction, this histidine abstracts the proton of the alcohol hydroxyl group while a hydride is transferred to the FAD N5, oxidizing the initial hydroxyl to a carbonyl and so, the alcohol to its respective aldehyde.

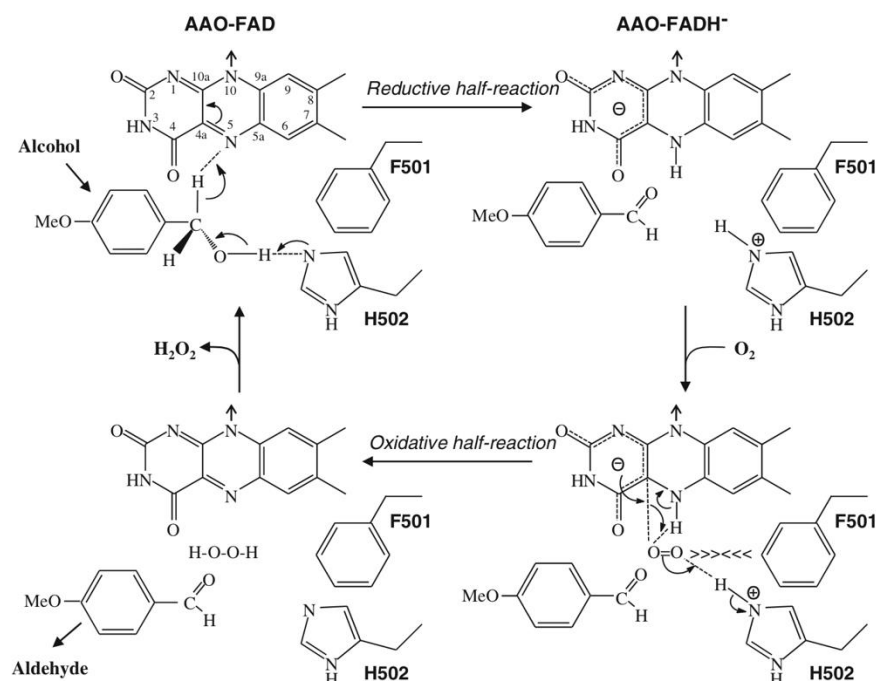


Figure 5. Scheme for Aryl-alcohol oxidase catalytic cycle as an example of a flavoprotein reaction mechanism (from Hernández-Ortega et al. 2012)

Sometimes, this can be further oxidized to a carboxylic acid. This can only occur if the substrate adopts the gem-diol structure through the hydration equilibrium (Figure 6). A second catalytic residue is present right next to the first one, mostly a histidine, or asparagine to a lesser extent. Its function seems to be related to the proper placement of the hydroxyl group for the proton transfer, by means of a hydrogen bond with the oxygen of the alcohol. In the oxidative half-reduction, the oxygen is reduced to hydrogen peroxide. But molecular oxygen is found in the triplet state as the most stable electronic conformation, or ground state, while FAD is in the singlet state. Since transitions from singlet to triplet are spin-forbidden, the oxygen molecule needs to be activated. The generally accepted mechanism for the oxygen activation consists of an electron transfer from the reduced flavin to the oxygen molecule, generating the superoxide anion. Then, it can form a transient C(4a)-hydroperoxyflavin or proceed with an outer-sphere second electron transfer (Romero et al. 2018). This last step seems to be system-dependent.

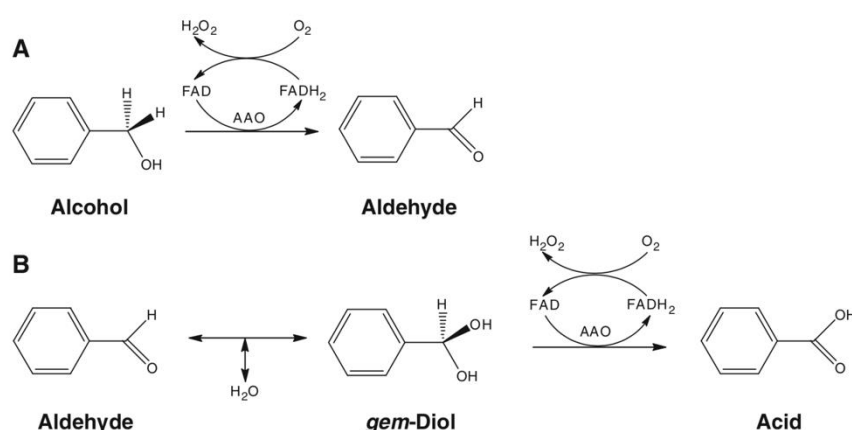


Figure 6. Scheme for reactions catalyzed by AAO. AAO oxidizes aromatic alcohols to the corresponding aldehydes (A) and it can also oxidize aromatic aldehydes to their corresponding carboxylic acids (B) (from Hernández-Ortega et al. 2012).

Interestingly, flavoproteins tend to perform better around neutral pH conditions, which corresponds to histidine's side chain pK_a (although this value can diverge depending on the residue environment). The reason is that the catalytic histidine must be able to either accept a proton from the substrate or release it in benefit of the final electron acceptor, in order to return to the catalytically active redox state.

Flavoproteins are valuable and cost-effective biocatalysts since they do not require expensive coenzymes, but just use molecular oxygen as a co-substrate. Considering their chemical versatility, flavins are involved in essential biochemical reactions in most organisms (Jortzik et al. 2014). For this reason, flavoproteins are potential targets of

pharmacological treatments. Their versatility along with other properties such as enantioselectivity or the potential *in situ* obtention of hydrogen peroxide as a sub-product, flavin-dependent biocatalyst have a huge potential in the industry of compound synthesis (Baker Dockrey and Narayan 2019).

2.2.1. Aryl-alcohol oxidase

Aryl-alcohol oxidase (AAO) is a flavoprotein that belongs to the glucose-methanol-choline (GMC) oxidase family. This enzyme has been found in *Pleurotus eryngii*, a mushroom known by its selective degradation of lignin in lignocellulosic materials (Sonnenberg et al. 2016). Scientists have shown that the role of this enzyme in the lignin degradation is to provide the hydrogen peroxide required by other ligninolytic (Ferreira, Hernandez-Ortega, and Herguedas 2009). Although it may accept other primary alcohols, the extracellular *p*-anisyl alcohol (ANI) is considered to be the natural substrate of this enzyme. Because of the ecological and biotechnological relevance of lignin degradation, this enzyme has been an object of intense study for the past few years. Thus, the role of several amino acids in the catalytic mechanism has been revealed.

This flavoprotein does not have a covalently-bound FAD, but the cofactor is deeply buried inside the protein. Two residues are considered catalytic: i) His502, which acts as a base extracting the proton of the hydroxyl group as previously explained and ii) H546, that interacts with the oxygen of the alcohol group for correct positioning of the substrate reaction (Hernández-Ortega et al. 2011). The surface of the protein is connected to the active site through a funnel-shaped channel, constricted by three residues (Tyr92, Phe397 and Phe501) whose side chains form a bottleneck. This regulates the access of substrates in the buried active site. Moreover, while the presence of an aromatic residue at positions Y92 and F501 has been demonstrated to be important for the oxidative half-reaction (Hernández-Ortega et al. 2011), it produces a steric hindrance when the substrate is substituted by a secondary alcohol (i.e. the substitution of one hydrogen by a methyl group) (Hernández-Ortega et al. 2012). The structure around the active site is illustrated in Figure 7.

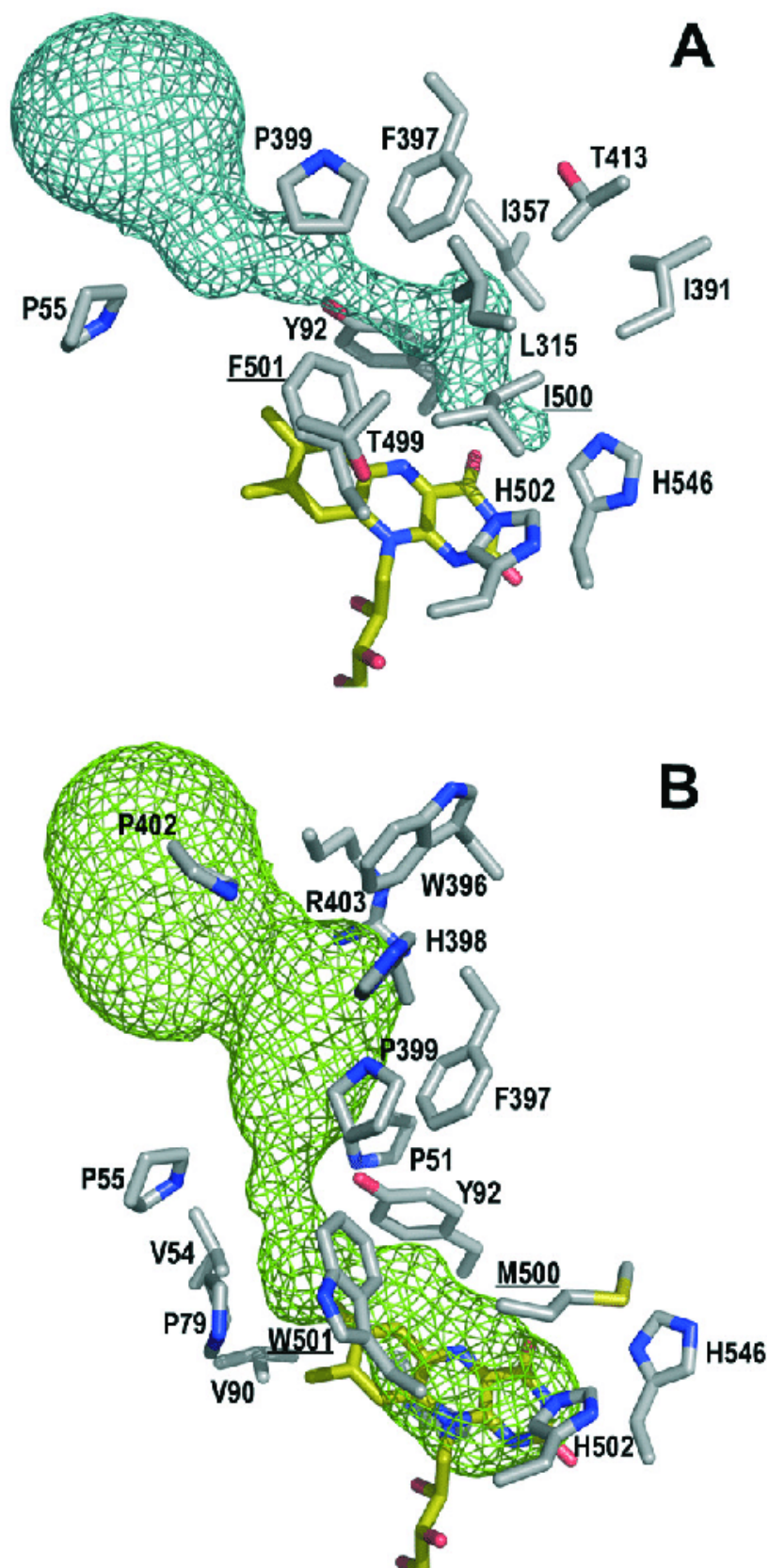


Figure 7. Detailed structure around the active site cavity of AAO (from Serrano et al. 2019).

2.2.2. Hydroxymethylfurfural Oxidase

In the pursuit of second-generation biofuels and biochemicals from lignocellulosic biomass, toxic inhibitors such as 5-hydroxymethylfurfural (HMF) limit their utility. In 2010, the metabolic pathway and HMF metabolizing bacteria were isolated (Koopman et al. 2010). From there, the enzyme 5-hydroxymethylfurfural oxidase (HMFO) was found from *Methylovorus sp.* strain MP688, a group of Gram-negative bacteria. This enzyme catalyses the oxidation of HMF to 2,5-furandicarboxylic acid (FDCA) in several steps: i) HMF is oxidized to the corresponding aldehyde generating 2,5-diformylfuran (DFF). Through spontaneous hydration to gem-diol (as previously mentioned), this compound is further oxidized to 5-formyl-2-furancarboxylic acid (FFCA). Once again, after hydration, it gets oxidized to the final product, FDCA (Dijkman et al. 2015). As in the AAO system, the FAD-containing protein holds the cofactor in the active site tightly, but not covalently bound. Here, His467 acts as a base, while the hydroxyl fixing role is carried by an asparagine, N511.

Although the full oxidation of HMF to FDCA catalysed by HMFO is possible, it is not very efficient. The overall catalytic efficiency improves by the introduction of W466F and V367R mutations, which increase the rate of the limiting step (FFCA oxidation) at expenses of a decrease in the catalytic rate of the first two oxidation (Martin et al. 2018; Dijkman et al. 2015; Pickl et al. 2018).

2.3. Laccases

Laccases are copper-containing oxidases (Morozova et al. 2007). These metal atoms can be found as divalent cations inside the enzyme, organized in two clusters: i) one t1-type Cu^{2+} (T1) and ii) three Cu^{2+} , atoms forming the so-called trinuclear cluster (TNC) (see Figure 8). At the same time, the TNC is formed by two t3-type (T3) and one t2-type (T2) Cu^{2+} . The position of these atoms inside the laccase is fixed by coordination with different residues. The TNC is solely bound to the imidazole side chains of several histidines, while the T1 copper has two histidines and one cysteine as equatorial complexing agents and isoleucine and an unspecific residue in the axial positions. This unspecific residue can be a methionine in laccases with low redox potentials, isoleucine or phenylalanine.

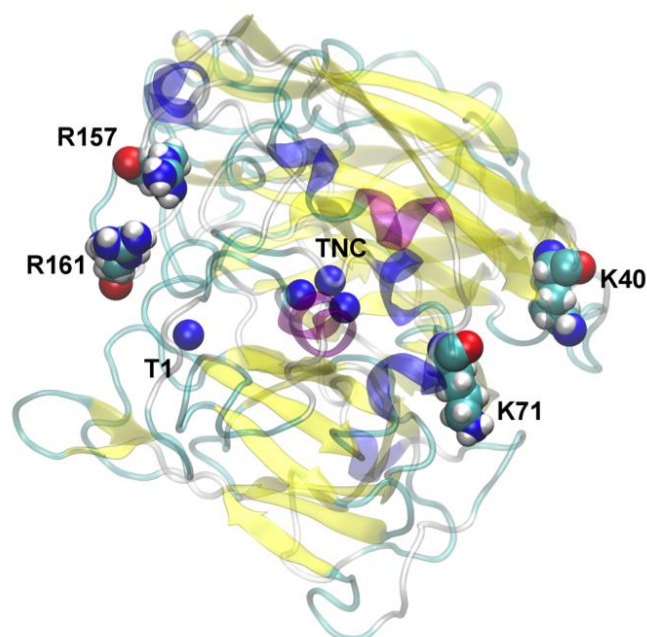


Figure 8. Laccase structure, where relevant ions and residues are highlighted (from Robert et al. 2017).

As an oxidase, the function of this enzyme is to catalyse one-electron oxidations. Due to an exposed superficial active site, laccases can perform over a wide range of compounds, being oxygen the final electron acceptor. When the substrate binds into the laccase's active site, and the complex is formed, one of the T1 equatorial histidines mediates the transference of one electron towards the Cu^{2+} , reducing it to Cu^+ . In some cases, such as the oxidation of phenols and amines at low pH, the electron transfer can be coupled to a proton transfer. In these situations, an aspartic acid will behave as a catalytic base. After the reduction of the T1, the electron tunnels through the equatorial cysteine and through two of the T3-coordinating histidines before reaching the TNC. Since O_2 is the final electron acceptor, which will be reduced to water, four electrons are needed to complete the catalytic cycle and thus, four substrates are required to fully reduce the enzyme and activate the oxygen reduction.

Due to their great versatility in oxidizing different substrates and the use of a cheap sacrificial electron acceptor, which furthermore produces water as a by-product, laccases have been classified as seductive for green chemistry (Riva 2006). Therefore, many engineering efforts have been dedicated to the improvement of laccases, by increasing their activity or oxidizing new compounds. In particular, these efforts have been focused on incrementing the redox potential of the T1 copper, since it was found that, for some substrates) the T1-substrate redox potential difference was correlated with the rate constant (k_{cat}) of the reaction (F. Xu 1996; Tadesse et al. 2008). However,

increased T1 redox potentials have not always been translated into better activities (Durão et al. 2006). Therefore, it is not clear whether if playing with the T1 redox potential is the best strategy to improve laccases since it also affects the internal electron transfer between the T1 and the TNC (which is already energetically hampered). Other strategies have been proposed for laccase improvement, that focus on enhancing the protein-ligand interaction. Mutations in the active site near the T1 allow the substrate to adopt a more buried catalytic position, providing a more favourable electrostatic environment for the electron transfer to happen (Monza et al. 2015). However, it is difficult to achieve the oxidation of otherwise inactive compounds, with high redox potentials, by tuning its binding. To solve this problem, the use of mediators has been shown to be a more successful approach (Riva 2006).

Mediators are molecules that add an intermediate step in the oxidation of a molecule. The mediator, which reacts directly with the enzyme, is oxidized by the T1 copper and, can then proceed to remove one electron from other molecules. Interestingly, if the redox potential of the mediator oxidized form is larger than that of the T1, the system could potentially oxidize a wider range of substrates. The full reaction mechanism including the mediator is represented in Figure 9.

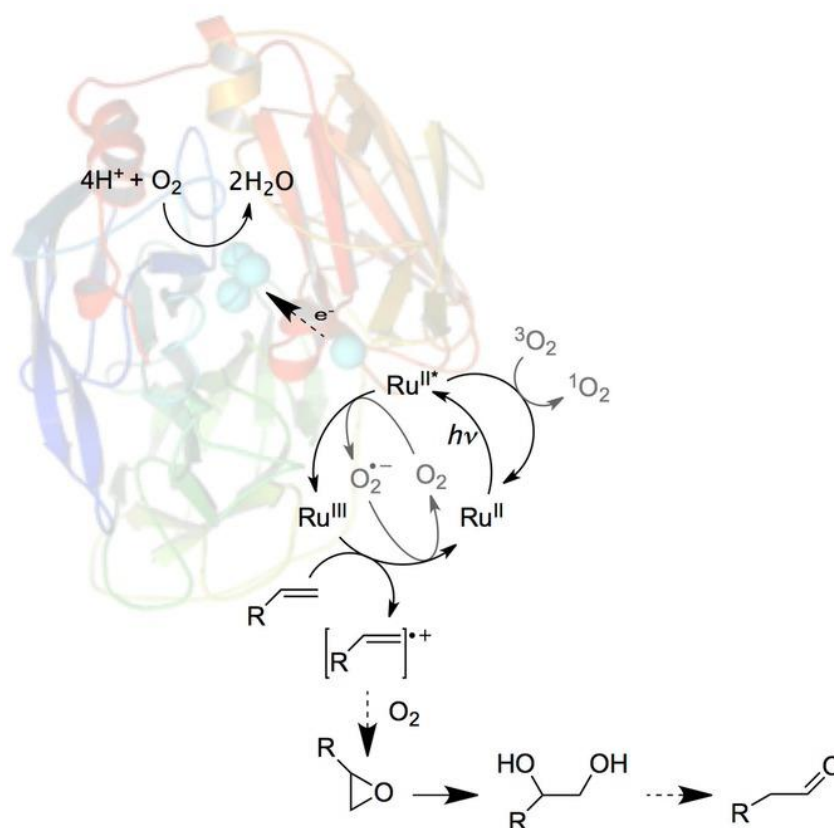


Figure 9. Full reaction mechanism mediated by a ruthenium photosensitizer (from Robert et al. 2017).

As in the case of flavoproteins, the number of proposed applications is larger than the actual number of laccases being used in the industry (Marques et al. 2018; Kunamneni et al. 2008). Textile, paper and food industries have laccases consolidated in their processes, while more candidates are still exploring their possible commercialization, such as organic synthesis, pharmaceutical or nano-biotechnological industries.

2.3.1. The use of a photosensitizer with laccases

Tron and co-workers investigated the association of laccases with ruthenium-polypyridyl photosensitizers for epoxidation. They showed that the cooperativity of a laccase-tris(bipyridine)ruthenium(II) ($[\text{Ru}(\text{bpy})_3]^{2+}$) system was able to oxidize compounds considered prohibitive because of their redox potential (Schneider et al. 2015). The main asset of this mediator is that, after photoexcitation with visible light, a very strong oxidant is obtained, with a high redox potential (see redox potentials in Table 2). This increased the range of substrates that can be oxidized. When visible light is shed, the ruthenium complex gets excited, leading to an incredibly fast electron transfer towards the T1 copper (Simaan et al. 2011). However, bi-molecular systems are limited by diffusion, a problem that can be solved by grafting $[\text{Ru}(\text{bpy})_3]^{2+}$ to the laccase. To do so, one bipyridine ligand has to be fused with an imidazole motif and benzaldehyde.

Tagging methods are generally based on the chemical properties of the residue side chains. For instance, lysine is an amino acid with a reactive function. The problem is that they tend to be abundant, difficulting the specificity of the grafting process. For this reason, LAC3 was found to be the optimal protein for this study, since only two lysines are present in the native sequence obtained from the fungus *Trametes sp.* C30 (Klonowska et al. 2005).

LAC3 is a fungal laccase involved in lignin degradation and detoxification of lignin-derived products (Klonowska et al. 2005). This protein has only two lysine residues present among the 501 residues, which are close to the TNC side. Therefore, this enzyme can easily be engineered to remove these lysines (obtaining a laccase with no lysines, K0), and then designing variants with just one reactive amino acid to precisely bind covalently the ruthenium complex. These variants with only one lysine will be denominated as UNIK variants. Other interesting residues that are relevant for this thesis are R157 and R161, positions where lysines are sometimes found in laccases, and that are closer to the T1 copper (rather than lysines originally present in the native LAC3).

The use of the LAC3/[Ru(bpy)₃]²⁺ system as a powerful oxidase has, nonetheless, trade-offs. The high redox potential has a drawback, which is a significant back-transfer rate, even faster than the internal electron transfer between the T1 and the TNC (the estimations of these values will further be discussed). This means that, after photoexcitation and T1 reduction, the oxidized ruthenium complex would recover the electron before it can reach the TNC. By modifying the interaction between the laccase and the photosensitizer, it could be possible to overcome this limitation.

Table 2. Relevant redox potentials in this thesis (Juris and Moggi 1981; Simaan et al. 2011).

Redox couple	E° (V)
$\text{Cu}^{2+}(\text{T1, LAC3}) + \text{e}^- \rightarrow \text{Cu}^+(\text{T1, LAC3})$	0.68
$[\text{Ru}(\text{bpy})_3]^{3+} + \text{e}^- \rightarrow [\text{Ru}(\text{bpy})_3]^{2+}$	1.29
$[\text{Ru}(\text{bpy})_3]^{3+} + \text{e}^- \rightarrow [\text{Ru}(\text{bpy})_3]^{2+*}$	-0.81

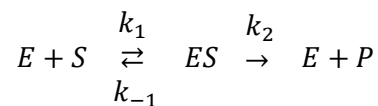
3. Theory

In this chapter, the theoretical background of the methods used in this thesis is exposed. First, a brief introduction to the general theory behind protein kinetics and protein-ligand interaction. This is followed by the computational modelling methods used.

3.1. Protein related theory

3.1.1. Enzyme kinetics

In general, we assume that the enzymes follow the Michaelis-Menten kinetic model (Michaelis et al. 2011). Therefore, we can assume a single-substrate enzymatic model reaction with the following equation:



Equation 1. Reaction model of an enzymatic reaction.

This equation implies that a substrate (S) binds to an enzyme (E) to form the Michaelis complex (ES) with a certain rate of formation (k_1). Then, the Michaelis complex can proceed irreversibly towards product formation, with a rate (k_2) or dissociate to the initial substrate and enzyme (k_{-1}). The rate of the global enzymatic reaction can be modelled by the Michaelis-Menten equation.

$$v = \frac{v_{max}[S]}{K_M + [S]} = \frac{k_{cat}[E][S]}{K_M + [S]}$$

Equation 2. Michaelis-Menten equation.

Here, the k_{cat} is the turnover number (the number of times a substrate is converted into a product per unit of enzyme and time). It is directly related to the maximum substrate to product conversion rate (v_{max}) when for a given enzyme concentration [E], the substrate saturation concentration [S] is reached. This implies that the Michaelis-Menten constant (K_M), defined as the substrate concentration needed to half of v_{max} , is negligible under these conditions. When this is experimentally reproduced in the lab, with an initial concentration of substrate and enzyme so that $[S] \gg [E]$, two different stages can be appreciated (Figure 10): i) the concentration of Michaelis complex ES increases over

time (pre-steady state) and ii) the concentration of the same complex remains constant over time (steady-state). In the pre-steady state approximation, when the product formation is slower than the dissociation (and thus, $k_2 \ll k_{-1}$), K_M coincides with the dissociation constant of the Michaelis complex ES.

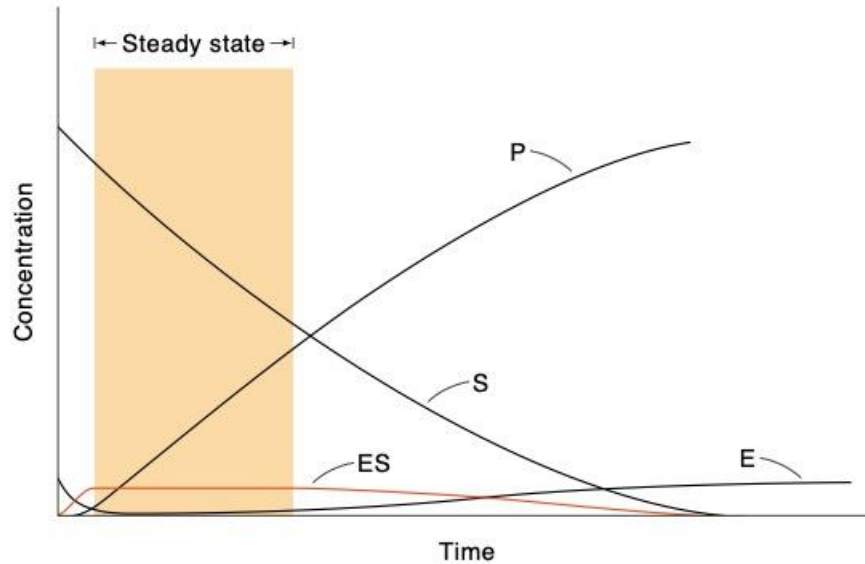


Figure 10. Plot of the concentrations of the various species versus time in a Michaelis-Menten model enzymatic reaction (modified from <http://www.uscibooks.com/changten.pdf>).

Therefore, when these conditions are fulfilled, K_M can be a measure of the enzyme-substrate affinity. However, under the steady-state approximation conditions, the product formation is much faster than the complex formation, ($k_2 \gg k_1$), and the concentration of the Michaelis complex tends to zero. In this framework, K_M has a direct dependency over k_{cat} .

$$K_M = \frac{k_{cat} + k_{-1}}{k_1}$$

Equation 3. K_M dependency with k_{cat} .

Nevertheless, these equations cannot always be applied straightforward. Some reactions can have more than one substrate (as explained in the flavoproteins section on chapter 2) or the product release can be so slow that it would have to be considered a separate stage in the kinetic equations as $EP \rightarrow E + P$. For these reasons, computational estimation of K_M is not trivial, since it can potentially depend on multiple rate constants.

On the other hand, k_{cat} can be calculated on simple reactions where its rate-limiting step (the slowest step) is known. If so, by modelling the transition state and the application of

the Eyring's equation (or Marcus' equation if an electron transfer is involved), k_{cat} could be obtained.

3.1.2. Electron transfer in proteins

The kinetics of a reaction involving electron transfer processes are governed by the Marcus theory (Marcus 1956). This theory determines that the electron transfer reaction rate depends on three parameters: i) the electron coupling, which is the tunnelling probability between electron donor and acceptor wavefunctions, ii) the driving force, the overall Gibbs free energy of the electron transfer and iii) the reorganization energy, which is the name that receives the energy cost of the geometry changes between reactants and products. These three parameters are present in the Marcus equation,

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

Equation 4. Marcus equation.

where \hbar is the reduced Planck's constant, K_B the Boltzmann's constant and T the temperature.

The order of magnitude of the pre-exponential factor of the Marcus equation is dictated by the magnitude of the electron coupling. If this parameter is too large, the Marcus theory cannot be applied. The electron coupling depends on the donor-acceptor distance and in the nature of the tunnelling events, which can be, in increasing order, through vacuum, H-bonds or covalent bonds. In proteins, and laccases in particular, electron transfer reactions involve relatively long distances, assuring the applicability of this theory (Tadesse et al. 2008).

In the exponential factor, the driving force depends on the redox potential difference between donor and acceptor. It represents the facility for the electron to go from the donor to the acceptor. Following the equation, the electron transfer rate maximum is given when the sum between the driving force and the reorganization energy is equal to zero.

As mentioned earlier, the reorganization energy is defined as the energy to restructure the system from the initial to the final coordinates, without considering the electron transfer. This can be split into two contributions: i) inner-sphere, which strictly depends

on the geometry of the donor and the acceptor and ii) the outer-sphere, that includes the rest of the system, typically other amino acids involved and water molecules.

3.2. Computational enzyme modelling

The complexity of enzymes, due to the large number of variables to be considered, makes their study difficult. For this reason, there are many methods, involving different levels of theory. Here, the methods used in this thesis are introduced.

3.2.1. Quantum Mechanics

In order to study properties of an enzyme that require an explicit treatment of the electronic degrees of freedom, quantum mechanics-based methods offer the most accurate description of a system at the expense of increased computational costs. Considering a stationary system where the Hamiltonian operator does not depend explicitly on time, the time-independent Schrödinger equation can be applied to obtain the energy and the wavefunction of the system:

$$H\Psi = (T_N + T_e + V_{NN} + V_{eN} + V_{ee})\Psi = E\Psi$$

Equation 5. Time-independent Schrödinger equation.

where T and V are the kinetic and potential operators respectively, of the nuclei (N), electrons (e) and their interactions: nuclei-nuclei (NN), electron-nuclei (eN) and electron-electron (ee). As stated in the first postulate of quantum mechanics, the wavefunction contains all the information of the system (Atkins and Friedman 2011). Thus, the application of the operator of an observable can be applied to obtain its estimation value. In the time-independent Schrödinger equation, the Hamiltonian is the operator of the observable “energy”, which can be divided into the sum of the kinetic and potential energies. Due to the complexity of this equation, the position of the nuclei is considered fixed in relation to the electrons. This approximation, known as the Born-Oppenheimer approximation, can be applied because, due to the much higher mass of the nuclei, its movement is negligible on the electronic motion timescale. By removing the nuclear kinetic energy, we can isolate the electronic nuclear degrees of freedom from the Schrödinger equation, leading to the electronic Schrödinger equation (Szabo and Ostlund 2012).

$$H_e \Psi_i(r; R) = (T_e + V_{ee} + V_{eN} + V_{NN}) \Psi(r; R) = E_i \Psi_i(r; R)$$

Equation 6. Electronic Schrödinger equation.

In this equation, the coordinates are separated into electronic (r) and nuclear (R). However, for many-electron systems, this equation cannot be solved exactly and thus, more approximations are required. In order to briefly review some of these approximations, and to avoid classifying the density functional theory (DFT) into a group, three groups will be here explained: Hartree-Fock based (HF) ab initio methods, DFT and semi-empirical approaches.

Hartree-Fock based methods

To solve the time-independent Schrödinger equation of an N electron system, an initial guess of the wavefunction is needed. This wavefunction ansatz is built from the product of the mono-electronic wavefunctions or spin-orbitals (φ). To assure that the Pauli exclusion principle is satisfied, an anti-symmetrizer operator (A) is included. The resulting HF ansatz is known as the Slater determinant:

$$\Psi_{HF} = \sqrt{N!} A \prod_{i=1}^N \varphi_i(x_i) = (N!)^{-1/2} \sum_P (-1)^\pi P \prod_{i=1}^N \varphi_i(x_i)$$

Equation 7. Hartree-Fock ansatz.

where P is the permutation operator, N is the number of electrons in the system, π is the permutation's parity and x_i is the product of the spatial (r_i) and spin (σ_i) coordinates of the i -th electron. It is important to mention that spin-orbitals are constructed from the linear combination of atomic spin-orbitals (LCAO), mono-electronic functions centred on the nuclei of the atom. The reference atomic spin-orbitals are called basis set. The accuracy of the results depends directly on the number of atomic spin-orbitals and the quality of the basis set (Schlick 2010).

$$\varphi_i(x_i) = \sum c_{\mu i} \chi_{\mu}(x_i)$$

Equation 8. Spin-orbitals as combination of basis set.

From the Slater determinant, the Schrödinger equation can be solved by applying the variational principle, which states that the expectation value of the energy $E[\psi]$ cannot be lower than the ground state energy E_0 .

$$\delta E[\Psi] = 0; \quad E[\Psi] = \frac{(\Psi|\hat{H}|\Psi)}{(\Psi|\Psi)}; \quad E[\Psi] > E_0$$

Equation 9. Variational principle.

Skipping all the mathematical treatment and focusing only on the concepts, after minimizing and some algebraic manipulation, from equation basis-set, the following expression is obtained, already in matrix form:

$$FC = SC\varepsilon$$

Equation 10. Roothaan-Hall equation in matrix form.

where F is the Fock matrix, C is the coefficient matrix from the basis set equation, ε is the orbital energies matrix and S is the overlap matrix, that imposes the orthonormality of spin-orbitals. Since F depends on C, this equation cannot be solved by simply diagonalizing. Therefore, an iterative procedure called the self-consistent field approach is used, by which the energy is obtained by convergence regarding a predetermined threshold.

Using HF, there are different ways to treat electrons with opposite spin: i) occupying the same molecular orbital, with a different spin component (restricted, RHF), ii) using different molecular orbitals for electrons with opposite spin (unrestricted, UHF) or iii) using different molecular orbitals for paired and unpaired electrons (restricted-open, ROHF). However, the repulsion between electrons is not explicitly treated, leading to significant errors. This error is the correlation energy, the difference between the exact energy and the one obtained by the HF method (in a complete basis set). There exist methods that try to minimize this error by adding more Slater determinants to the HF ansatz, such as configuration interaction (CI) or coupled-cluster (CC). Others apply perturbation theory, like the Møller-Plesset (MP) method. Nevertheless, the computational costs of these methods are prohibitive in biomolecular simulations yet.

Density functional theory

In Density functional theory or DFT, widely used due to its compromise between accuracy and computational costs, the role of wavefunctions is replaced by the electron density (the probability of one electron to be in a certain location) (Koch and Holthausen 2015). The main advantage is that the electron density probability is limited to the three-dimensional spatial degrees of freedom (x,y,z) which are independent of the system size.

Also, it provides a direct bridge between theory and experiments since it can be measured by X-ray or electron diffraction techniques. Furthermore, DFT accounts for electronic correlation still keeping the computational cost at bay.

The Hohenberg-Kohn theorems settle the base of DFT. They relate to any electron-involving system moving under the presence of an external potential, and state that:

Theorem 1: The external potential (and hence the electronic energy) is a unique function of the ground state electron density.

Theorem 2: The functional that gives the energy of the ground state delivers the lowest energy only if the input density is the true ground state density. In other words, the ground-state electron density can be obtained variationally.

However, the universal functional is unknown and these theorems do not provide any protocol to calculate the ground-state electron densities. To overcome these limitations, the Kohn-Sham (KS) equations (Kohn and Sham 1965) can be applied, which express the electron density of a system as a linear combination of basis functions, borrowing the concept from HF. So, the universal functional can be written as:

$$E[\rho(r)] = T_S[\rho(r)] + E_H[\rho(r)] + E_{XC}[\rho(r)]$$

Equation 11. Universal functional according to the Kohn-Sham theory.

where $E[\rho(r)]$ is the spin density energy functional, $T[\rho(r)]$ is the uncorrelated kinetic energy functional, $E_H[\rho(r)]$ is the classical electrostatic energy functional of the non-interacting system and the term $E_{XC}[\rho(r)]$ corresponds to the exchange-correlation functional, which is the only unknown functional, and includes the corrections of the two previous functional to the exact solution. Therefore, there are several approximate functionals that have been developed trying to reproduce experimental results by parameterization. For this reason, this method is found in the limbo between *ab initio* and semi-empirical methods, while others even doubt about its consideration as a QM method.

Semi-empirical approaches

These methods have been increasing in popularity for biomolecular simulations in the past few years (Yilmazer and Korth 2015). Semi-empirical approaches introduce more

approximations to the HF method, decreasing the complexity of the Fock matrix. This is achieved by introducing empirical data, so the wave function interactions are not calculated explicitly but built from predefined parameters. Therefore, only valence electrons are treated quantumly. This approximation saves a lot of the simulation computational cost at the expense of accuracy, compared to HF methods.

In this thesis, QM methods have been mainly applied to calculate the atomic charge of small molecules, such as ligands and cofactors. When using classical methods, the atoms in proteins are already parameterized in the force field (see next section). The atomic charge, which is the amount of total charge assigned to a particular atom, can be assigned by dividing the wavefunction (Mulliken 1955) or the electron density (Bader 1994). Even though Mulliken charges are adequate in general terms, they are basis-set dependent. Another approach is electrostatic potential (ESP) fitting, which calculates the electrostatic potential at each point of a grid outside the molecule and then assigns the atomic charges of the atoms trying to reproduce the same potential.

3.2.2. Molecular Mechanics

In enzyme modelling, it is well known that protein flexibility plays a key role in the activity of the enzyme. Protein dynamics can affect the active site preorganization, the ligand binding and even the diffusion towards the active site or product release. Also, it has been shown that dynamic effects can hide relevant conformations (Hart et al. 2016), otherwise hidden in the crystal structure. The study of these processes becomes impossible for most QM methods, due to its high computational cost. For this reason, these large biomolecules are usually treated classically, meaning that the electronic degrees of freedom are not treated explicitly.

In molecular mechanics (MM), the potential energy of a system is fitted over a set of experimental or QM data (Leach 2001). Therefore, functional groups are assumed to keep their properties disregarding their environment: for instance, the side chain of an amino acid maintains its properties no matter what residues are surrounding it. This information is gathered in the so-called force fields (FF). We can differentiate in between i) all-atomistic models, where all the atoms are parameterized individually, or ii) coarse-grained methods, where atom groups with particular properties are considered a unique particle, reducing considerably the size of the system but also losing accuracy.

In general, FF includes two differentiated interactions, bonded and non-bonded interactions. Bonded interactions include stretching and bending energies, which are

described by harmonic functions, and torsional energies, written as Fourier series to ensure periodicity. On the other hand, the non-bonded interactions incorporate the electrostatic energies, which consider the coulombic interaction between atoms (represented as point charges) and van der Waals interactions. With a Lennard-Jones potential form, the van der Waals interactions acknowledge nuclear repulsion and dispersion forces. These energetic contributions are summarized in Figure 11.

$$\begin{aligned}
 U = & \sum_{i < j} \sum 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \\
 & + \sum_{i < j} \sum \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \\
 & + \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_0)^2 \\
 & + \sum_{\text{angles}} \frac{1}{2} k_a (\theta - \theta_0)^2 \\
 & + \sum_{\text{torsions}} k_\phi [1 + \cos(n\phi - \delta)]
 \end{aligned}$$

Figure 11: Scheme and equations of a general force field (from Anwar and Zahn 2017).

From these general formulas, different force fields may have different parameters or even include other terms, like the variable use of improper dihedrals, to keep the chirality and planarity of certain systems (Guvench and MacKerell 2008). Since enzymatic systems are typically simulated in an aqueous environment, explicit waters can be modelled through specific FF. In order to reproduce the bulk behaviour of the solvent, the number of water molecules included in the simulations would be too large, unnecessarily increasing the computational costs. To circumvent this issue, periodic boundary conditions (PBC) can be used by confining the system in a box that behaves as a unit cell, and it is replicated throughout the three-dimensional space by rigid translations. In the end, an infinite lattice is obtained. This is depicted in Figure 12. In order to avoid errors due to the PBC implementation, a cut-off radius is defined in a way that an atom can only interact with the closest images of the remaining atoms, following the minimum-image convention (Figure 12).

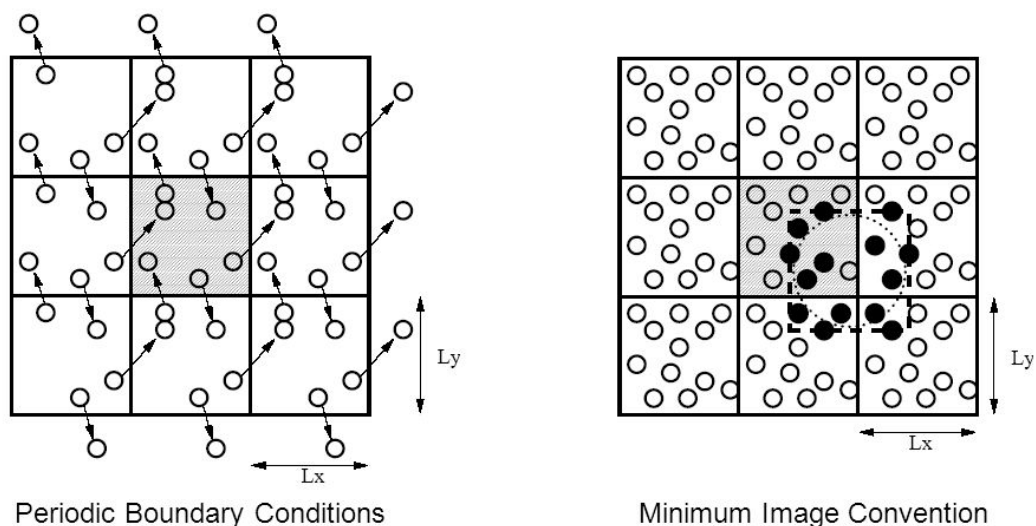


Figure 12. Graphical representation of periodic boundary conditions and minimum image convention (modified from Sumanth et al. 2005).

Alternatively, the effect of solvation in the free energy of the system can be considered with the use of implicit solvents (Cramer and Truhlar 1999). Even though this method has an important drawback in modelling compared to the explicit solvent, which is the inability to form hydrogen bonds, dealing with less computationally expensive systems can be useful for large macromolecular simulations. Also, not dealing with the explicit solvent molecules can be useful in some methods, such as the Monte Carlo-based PELE, which is explained in section 3.2.8. Implicit solvents have three contributions: electrostatics, van der Waals and cavitation, which considers the free energy penalty of creating a cavity in the solvent bulk. The last two contributions can simply be expressed as a linear function of the surface exposed to the solvent, known as the solvent-accessible surface area (SASA). However, the electrostatic term is more complicated to calculate. The Poisson-Boltzmann method would be the computationally expensive option, with increased accuracy, while the application of the generalized Born equation is faster.

Disregarding the specific FF to be used, it must be transferable in order to be general. This does not mean that they have to be universal (that can be applied to any system), but the parameters have to be transferable. For instance, the parameterization of an amino acid must be transferable to the same residue in different positions of the system, and also to different systems. For this reason, parameters have to be fixed. Therefore, by using a general force field like the ones introduced here, it is not possible to model reactivity, since bonds cannot be formed or broken. Moreover, polarization effects cannot be considered either. However, there are alternatives to overcome these limitations (Baker 2015; Kamerlin and Warshel 2011).

3.2.3. Quantum Mechanics / Molecular Mechanics

When the chemical reactivity of an enzyme is being modelled, the two methods exposed in the previous sections present some deficiencies. The most obvious one refers to what it was established at the end of the previous section regarding the formation or breakage of a bond, it simply cannot be modelled classically. The use of QM methods to reproduce reaction mechanisms has been successfully used in many cases, but in many others, the isolation of the atoms included in the QM simulation was leading to wrong solutions, since the influence of the environment was not being considered. The inclusion of more elements in the simulation could solve the problem, but computational costs could become prohibitive. Hence, the introduction of hybrid QM/MM methods (Warshel and Levitt 1976) provided a smart way to deal with the reactivity of large systems like enzymes. Its approach is very simple: the part involved in the reaction is modelled with QM, while the remaining non-reactive part is described by MM (Figure 13).

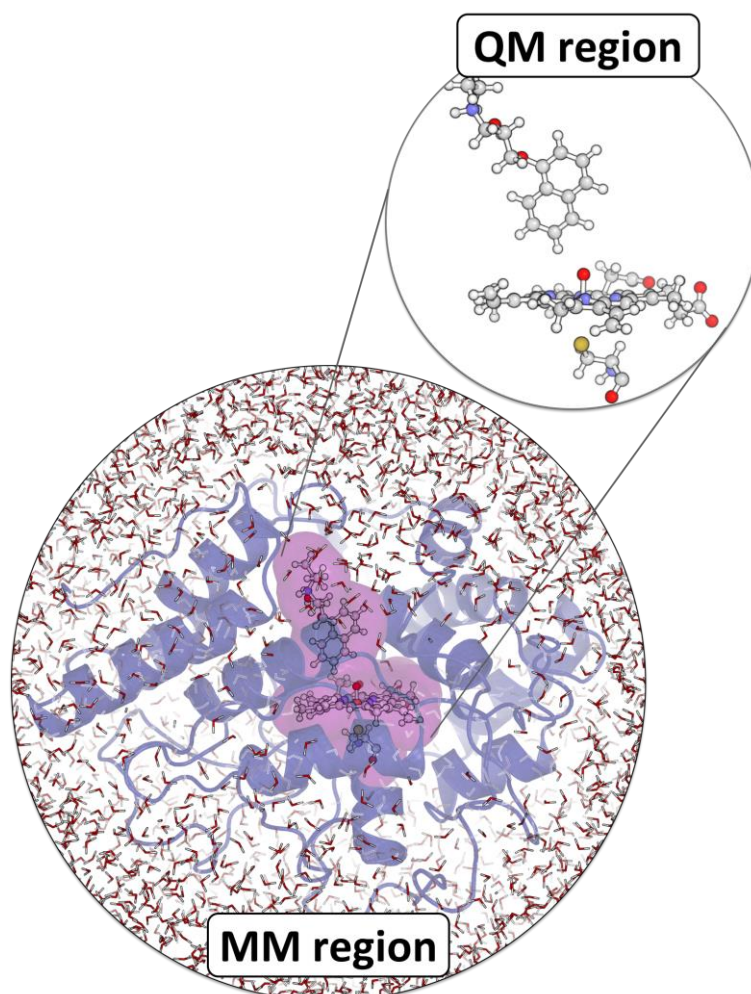


Figure 13. Schematic description of a QM/MM calculation (adapted from Romero-Rivera et al., 2017).

The most challenging part of QM/MM is how to deal with the interaction between the QM and MM regions. Considering an additive scheme, the energy of the system can be considered the sum of the QM and MM regions, plus a new term to consider the interaction between both regions (QM/MM).

In order to determine the $E_{\text{QM/MM}}$ contribution, from this point two possible scenarios can be presented to us, depending on whether the QM and MM regions are covalently bound or not. The latter would be the easiest case: if no covalent bonds are involved in between regions, van der Waals interactions are classically modelled by means of a Lennard-Jones potential, while the electrostatic part is calculated by enabling the polarization effects of the MM point charges to the QM electron density. However, when covalent bonds are present in the QM/MM interface, an approach to deal with these bonds is needed. Several methods have been suggested, such as the use of link atoms or frozen orbitals, but in this thesis, dangling bonds are capped by introducing a dummy atom (typically hydrogen) in the bond between the QM and MM atoms. The covalent energetic terms involving atoms from both QM and MM regions are treated classically. On the electrostatic part, atomic charges of the frontier MM atoms are replaced by spherical gaussian charges and redistributed in order to avoid over-polarization.

3.2.4. Geometry optimization

Up to now, different methods have been explained to quantify the energy of a system, from the accurate description of relatively small systems with QM, to the modelling of large systems by means of MM, and how they can be combined. However, these methods only calculate the energy of a given conformational structure. If this calculation was extended to all the possible conformations, the energy of this system could be represented as a function of its degrees of freedom, which is known as the potential energy surface (PES) of the system. From thermodynamics, the partition function is defined as:

$$z = \sum g_i e^{\left(-\frac{E_i}{kT}\right)}$$

Equation 12. Partition function.

Provided that the PES is known, from the partition function it can be determined the fraction of molecules in the energetic state E_i (Ott, Bevan Ott, and Boerio-Goates 2000). Moreover, it can be deduced that the conformations with the lowest energies will contribute the most to the partition function. These low energy values can be found in

the local minima of the PES. For this reason, conformations corresponding to energy wells in the PES are considered as highly representative of the system and therefore, interesting to model.

$$\frac{n_i}{N} = \frac{g_i e^{-\frac{E_i}{kT}}}{z}$$

Equation 13. Number of molecules in the energetic state E_i .

The energy of a system can be minimized numerically through an interactive geometry optimization, where the structure of the system is slightly modified at each step until a convergence criterion is reached. This is, typically, when the root mean square deviation (RMSD) of the energy gradient is under a certain threshold. These structural modifications are not random but guided by methods like steepest descent or the conjugate gradient method. After the energy minimization, the optimized structure is obtained. However, it cannot be guaranteed that this geometry corresponds to an energy minimum. It can only be assured that it landed on a stationary point of the PES and that the energy of the final geometry is lower than the initial one (so it cannot be an energy maximum). For a full description of the stationary point, the diagonalization of the Hessian matrix is required (a square matrix of second-order partial derivatives of the PES). If there are no imaginary frequencies, the optimized structure is considered to be in an energy minimum. Otherwise, it corresponds to a saddle point, with a special mention for the cases where only one imaginary frequency is found, which represents a transition state.

3.2.5. Molecular dynamics

The previous sections showed that i) energy calculation methods (QM, MM or QM/MM) provided points of the PES and ii) through geometry optimization, stationary points can be found, in a local minimum exploration of the PES. Here (and the following sections), methods that explore the PES outside the local minimum will be introduced.

Molecular dynamics simulate the evolution of the system geometry over time, following Newton's second law of motion. This states that the acceleration of a system depends on the net force acting on that system and its mass. Since the force can be expressed as the gradient of the potential energy, the gradient of the PES can be evaluated to calculate the forces acting on each particle of a system and propagate it over time:

$$a_i(t) = -\nabla_i U/m_i$$

Equation 14. Newton's second law of motion as a function of the potential energy gradient.

where $a_i(t)$ is the acceleration of the particle i at time t , m_i corresponds to its mass and $\nabla_i U$ is the gradient of the potential energy. Therefore, given an initial set of positions and velocities for every atom, the new positions and velocities can be calculated after each time step δt , typically set to 2 fs considering that bonds involving hydrogen atoms are constrained. The general scheme of the MD algorithm can be seen in Figure 14.

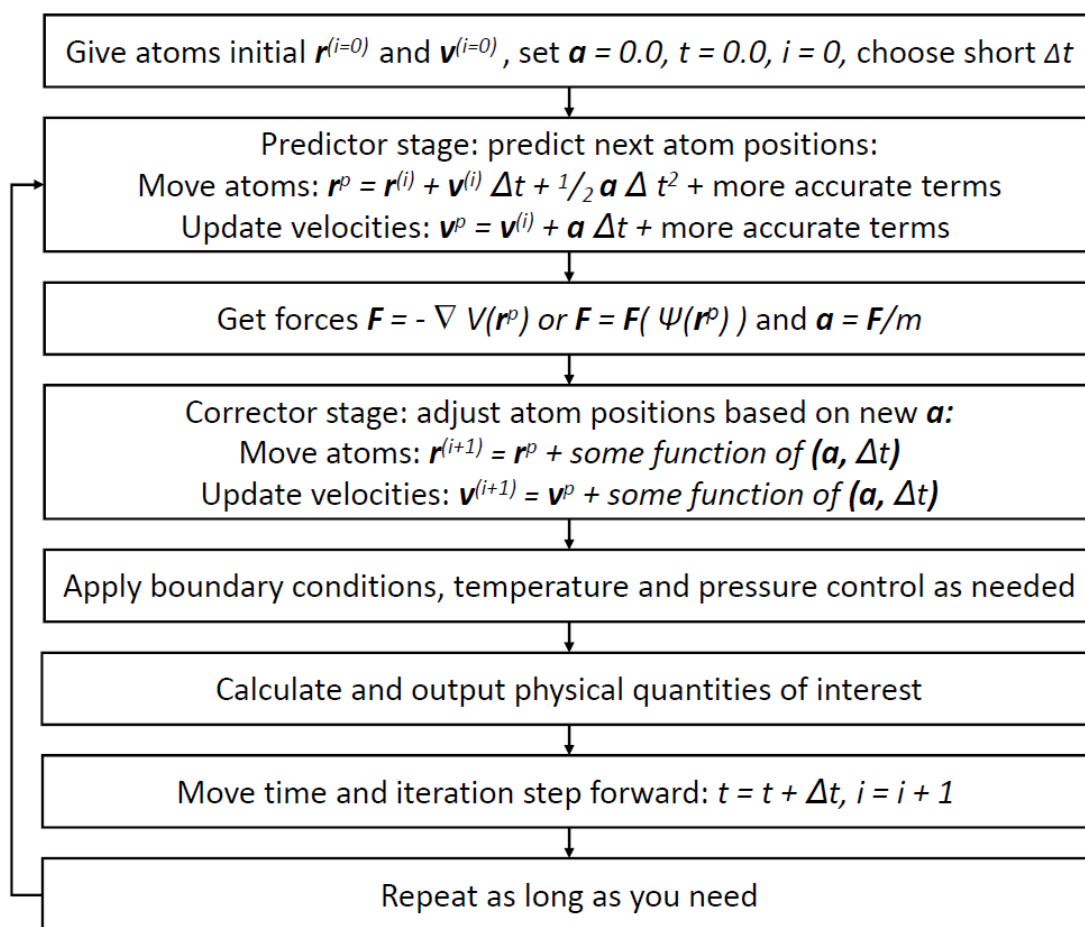


Figure 14. Basic scheme of the MD algorithm.

For protein studies, energies are generally obtained by means of MM, so the dynamics of the protein can be sampled at the nanosecond (ns) to millisecond (ms) timescales. In order to reproduce the experimental conditions, MD simulations usually simulated under an NPT ensemble, in which the number of particles, pressure and temperature of the system should remain constant during the simulation. This is achieved by the application

of a barostat and a thermostat, which is equivalent to rescaling coordinates and velocities.

Nowadays, it is still computationally expensive to run very long all-atom MD simulations, on timescales over 10 μ s. Although this problem can be tackled by technological advances, such as the use of graphical processing units (GPUs) or dedicated computers built specifically for MD simulations like the Anton machine (Shaw 2009), there are also algorithms that accelerate MDs. In this thesis, a particular method to do so has been used: metadynamics (Laio and Parrinello 2002). This method describes the free energy surface (FES), which is a Boltzmann weighted PES, of a chosen reduced set of collective variables (CV). The selection of a good set of CV is the key step: it has to be a representative variable, able to differentiate between the states of interest, and for this reason, no more than 2 or 3 CVs should be used. Then, a bias potential (a gaussian function of a CV) is added to the system to sample other values of a specific CV, modifying the FES. In the end, all the energy minima will be filled, resulting in a flat FES, where all the values of the CV would be considered explored and the simulation converged. At this point, the sum of all the Gaussian functions added for a CV is the inverse representation of the FES.

3.2.6. Monte Carlo

Monte Carlo (MC) based methods are presented as an alternative to MD simulations. Their differences rely on how the initial structures evolve. While, as explained in the previous section, MD follows Newton's second law of motion, new configurations are generated by a stochastic approach in MC. Therefore, the dependency on time is lost. In a Metropolis MC, a random move generates a new structure, that will be accepted or rejected following the Metropolis criterion. This is defined in the following equation,

$$W_{A \rightarrow B} = \begin{cases} 1, & \Delta E < 0 \\ e^{-\frac{\Delta E}{K_B T}} > z, & \Delta E > 0 \end{cases}$$

Equation 14. Metropolis criterion.

where $W_{A \rightarrow B}$ is the acceptance probability of going from state A to B, ΔE is the potential energy difference between states A and B, K_B is the Boltzmann constant and T reflects the temperature. As stated in equation 14, this means that a move will be directly accepted if the energy of the generated structure is lower than the initial one, otherwise,

a random number is generated. If this random number is lower than $e^{-\beta\Delta E}$ (where β is the inverse of the temperature times the Boltzmann constant) the perturbation will be accepted. Diversely, it is rejected. In order words, movements that increase the energy a little bit have more chances to be accepted than those increasing the energy a lot.

By removing the restriction of the Newton's equations of motions, MC methods offer cheaper computational costs and exceptional flexibility of the system with non-physical moves. These, together with the capability of accepting conformations with higher energy, according to the Metropolis criterion, make MC methods an interesting alternative to MD. On the contrary, MC methods lose the perspective of time and the physical reality of its movements and are usually limited to the use of implicit solvents.

3.2.7. Docking

Molecular docking techniques can be used when the goal is, instead of exploring the PES, to find the most stable conformation of a protein-ligand system. Generally, a confined search space is defined in which different orientations of the ligand are generated. The free energy of the system is then evaluated and ranked by means of a scoring function. The scoring function can be i) knowledge-based, using statistical factors from a structural database, ii) empirical, where the score depends on the nature of the contacts, iii) physics-based, if the energy is calculated using a FF or iv) derived from machine learning approached. There is a wide range of docking applications. For pharmaceutical studies, it is a very useful tool to find potential binders, by the evaluation of thousands of ligands against the target protein, in the so-called virtual screening (Kontoyianni 2017). On the contrary, in ensemble docking, the ligand is docked against multiple conformations of the same protein (obtained from an MD simulation, for example) trying to include the effect of the protein flexibility (Huang and Zou 2007). Precisely, this is one of the main drawbacks of this technique. Even though there are docking methods that sample the side chains of the active site, named induced-fit docking (M. Xu and Lill 2013), it increases considerably the simulating times, losing one of the main advantages of docking approaches.

3.2.8. Protein Energy Landscape Exploration algorithm

The main software used in this thesis, PELE (Protein Energy Landscape Exploration) is an MC-based in-house software (Borrelli et al. 2005; Madadkar-Sobhani and Guallar 2013) that aims to sample the protein-ligand conformational space. It combines protein

structure prediction techniques with random perturbations of the system. The idea behind this method follows the following scheme:

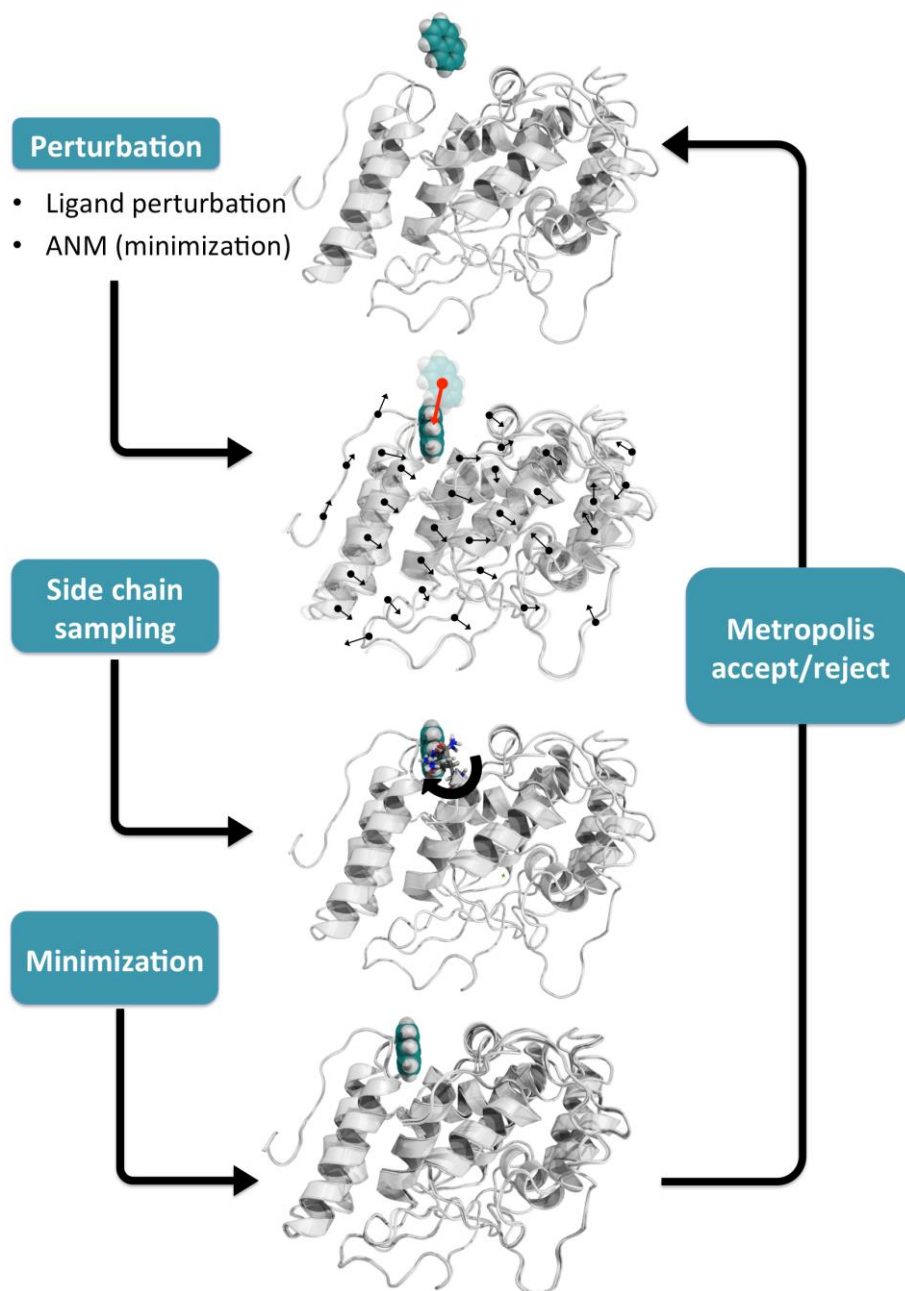


Figure 15. Graphical representation of a PELE step.

Given an initial protein-ligand system, at each step the ligand is randomly perturbed (rotated and translated) while the backbone of the protein is displaced through a linear combination of low-frequency normal modes. These are calculated with an anisotropic network model (ANM), that represents the protein as a network of elastic forces that connects the alpha carbons (Atilgan et al. 2001). Alternatively, modes provided by a principal component analysis could be provided. The vectors obtained determine the

movement of the protein backbone. The perturbation phase is followed by the relaxation step, consisting of a side chain prediction (Jacobson et al. 2002) and all-atom energy minimization. The former consists in selecting the best side-chain for each amino acid, from a discrete set of possible dihedral angle conformers (or rotamers), obtained from PDB statistics. The energy minimization is performed by a truncated Newton algorithm, using the Optimal Potential for Liquid Simulations (OPLS)-2005 FF and a generalized Born implicit solvent model for the solvation free energy calculation. After the system perturbation and relaxation, a new conformation is generated, which will be accepted or rejected according to the Metropolis criterion.

This algorithm is highly parallelizable, since each processor can produce a stochastic trajectory. Plus, in order to enhance the sampling, and benefiting from the inherent MC property that each step does not have memory from the previous ones, different stochastic trajectories can communicate to each other, in order to favour the sampling towards a goal (which can be, if known, the active site of a protein). Moreover, the MC sampling can be combined with an adaptive reinforcement learning procedure (Lecina, Gilabert, and Guallar 2017), for an increased performance in mapping the protein energy landscape.

Compared to the previously mentioned methods, PELE can performed a completely flexible protein-ligand docking (Borrelli, Cossins, and Guallar 2010; Cossins, Hosseini, and Guallar 2012) with moderate computational costs. Even though it was initially designed for mapping the ligand diffusion from and towards the active site (Borrelli et al. 2005), over the years it has shown that it can also be used to perform a blind active site search (without previous knowledge of its location) (Hosseini et al. 2015), or FES mapping with Markov state models (Takahashi, Gil, and Guallar 2014).

II. OBJECTIVES

The main objective of this thesis is to aim for improved enzymes that could be potentially used for industries and, in particular, for lignin degradation and valorisation. From a computational point of view, this includes providing new information to be tested in a lab, by suggesting positions where it would be interesting to perform saturated mutagenesis experiments, proposing punctual and specific mutations or even to describe the role of certain amino acids.

Therefore, the general objective of this thesis can be summarized as:

- Improvement of enzyme activity related to the degradation of lignin, mainly flavoproteins and laccases.

More specifically, this main objective is divided into the following sub-objectives:

1. Design AAO variants to enhance the oxidation of secondary benzyl-alcohols.
2. Study the full oxidation of HMF by means of flavoproteins.
3. Rationalize how laccases function with the photosensitizer and design improved variants to be tested in the lab.

III. RESULTS

1. Flavoproteins

1.1. Article I

Multiple implications of an active site phenylalanine in the catalysis of aryl-alcohol oxidase

Juan Carro¹, Pep Amengual-Rigo², Ferran Sancho², Milagros Medina³, Victor Guallar^{2,4}, Patricia Ferreira³ & Angel T. Martínez¹

¹ Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040, Madrid, Spain.

² Barcelona Supercomputing Center, Jordi Girona 31, E-08034, Barcelona, Spain.

³ Department of Biochemistry and Cellular and Molecular Biology, and BIFI, University of Zaragoza, E-50009, Zaragoza, Spain.

⁴ Icrea, Passeig Lluís Companys 23, E-08010, Barcelona, Spain.

Correspondence and requests for materials should be addressed to V.G. (email: victor.guallar@bsc.es) or P.F. (email: ferreira@unizar.es) or A.T.M (email: atmartinez@cib.csic.es).

SCIENTIFIC REPORTS

OPEN Multiple implications of an active site phenylalanine in the catalysis of aryl-alcohol oxidase

Received: 5 February 2018
Accepted: 11 May 2018
Published online: 25 May 2018

Juan Carro¹, Pep Amengual-Rigo², Ferran Sancho², Milagros Medina³, Victor Guallar^{2,4}, Patricia Ferreira³ & Angel T. Martínez¹

Aryl-alcohol oxidase (AAO) has demonstrated to be an enzyme with a bright future ahead due to its biotechnological potential in deracemisation of chiral compounds, production of bioplastic precursors and other reactions of interest. Expanding our understanding on the AAO reaction mechanisms, through the investigation of its structure-function relationships, is crucial for its exploitation as an industrial biocatalyst. In this regard, previous computational studies suggested an active role for AAO Phe397 at the active-site entrance. This residue is located in a loop that partially covers the access to the cofactor forming a bottleneck together with two other aromatic residues. Kinetic and affinity spectroscopic studies, complemented with computational simulations using the recently developed adaptive-PELE technology, reveal that the Phe397 residue is important for product release and to help the substrates attain a catalytically relevant position within the active-site cavity. Moreover, removal of aromaticity at the 397 position impairs the oxygen-reduction activity of the enzyme. Experimental and computational findings agree very well in the timing of product release from AAO, and the simulations help to understand the experimental results. This highlights the potential of adaptive-PELE to provide answers to the questions raised by the empirical results in the study of enzyme mechanisms.

Elucidation of structure-function relationships is of key importance in the study of enzyme catalysis. Unveiling the mechanisms that lie behind the properties of a biocatalyst paves the way to widen its biotechnological applicability by means of enzyme engineering¹. In this regard, aryl-alcohol oxidase (AAO) is an enzyme of biotechnological interest from the glucose-methanol-choline oxidase/dehydrogenase (GMC) superfamily. AAO catalyzes the oxidation of primary benzyl alcohols into their aldehyde counterparts using atmospheric O₂ as co-substrate and yielding H₂O₂ as co-product^{2–4}. In this way, it has shown its potential for the production of flavours and aromas, deracemisation of alcohol mixtures and the synthesis of precursors for the manufacturing of renewable polyesters⁵.

The ecophysiological role of AAO would be the supply of H₂O₂ to either ligninolytic peroxidases—which it acts synergistically with—or to trigger Fenton reactions during natural decay of lignocellulosic materials^{5,6}. H₂O₂ production by AAO involves the redox-cycling of *p*-anisaldehyde⁷, a metabolite of AAO-producing species^{8,9}. Its natural role as H₂O₂ producer may be exploited for biotechnological purposes by the development of enzyme cascades in which AAO and H₂O₂-consuming enzymes (peroxidases and peroxygenases) act concertedly. In recent years, the repertoire of AAO substrates has been enlarged by the discovery of new molecules the enzyme can oxidise, such as 5-hydroxymethylfurfural, 5-methoxymethylfurfural and their partially oxidised derivatives^{10,11}, and secondary benzylic alcohols. Regarding the latter substrates, AAO stereoselectivity and activity was improved by means of computer-guided rational design that facilitated a more appropriate positioning of benzylic alcohols in the active site thanks to the removal of the Phe501 side chain¹². AAO presents other potential applications in lignocellulose transformation and production of flavours and aromas⁵. Altogether, this variety of bioconversions renders AAO an enzyme with a bright future in biocatalysis.

The structure-function relationships of the model AAO from *Pleurotus eryngii* have been extensively investigated, and the roles of several residues important for catalysis have been elucidated^{13–17}. The catalytic pocket of *P. eryngii* AAO, located near the flavin moiety of its FAD cofactor, is shielded from the outer environment by a triad

¹Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040, Madrid, Spain. ²Barcelona Supercomputing Center, Jordi Girona 31, E-08034, Barcelona, Spain. ³Department of Biochemistry and Cellular and Molecular Biology, and BIFI, University of Zaragoza, E-50009, Zaragoza, Spain. ⁴ICREA, Passeig Lluís Companys 23, E-08010, Barcelona, Spain. Correspondence and requests for materials should be addressed to V.G. (email: victor.guallar@bsc.es) or P.F. (email: ferreira@unizar.es) or A.T.M. (email: atmartinez@cib.csic.es)

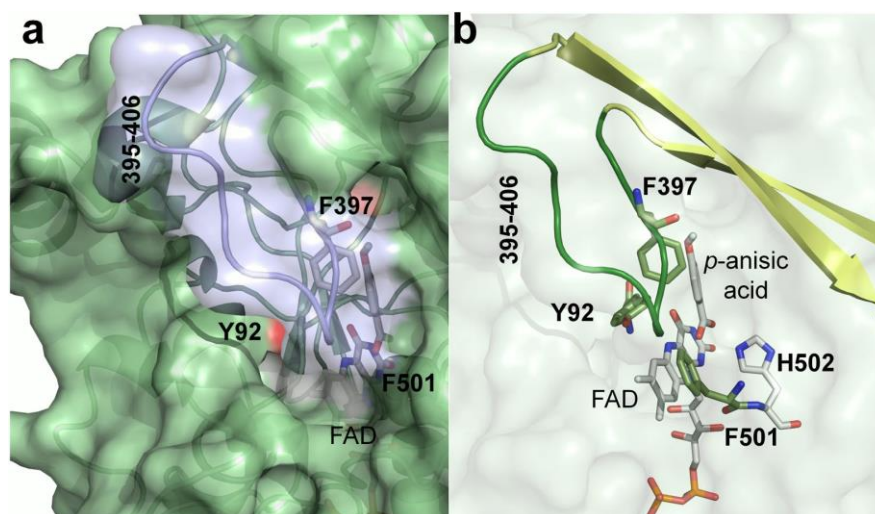


Figure 1. Surface loop and aromatic residues limiting access to the AAO FAD. (a) Area of channel opening on the AAO surface, with the 395–406 loop in transparent light blue. (b) Semitransparent surface showing the aromatic gate-keeping Tyr92, Phe397 and Phe501 aromatic residues (sticks in olive green), loop (green) and its contiguous β -strand residues (pale yellow cartoon), the catalytic His502 (grey) and FAD (grey), with one *p*-anisic acid molecule at the active site. From AAO:*p*-anisic acid structure (PDB 5OC1)¹⁷.

of aromatic residues —Tyr92, Phe501 and Phe397— that form a hydrophobic bottleneck affecting alcohol and oxygen migrations into AAO active-site^{18,19}. The roles of Tyr92 and Phe501 have been unveiled and are involved in: (i) establishing aromatic stacking interactions that guide the alcohol substrate to a catalytically competent configuration¹³; and (ii) compressing the active site to promote the reactivity with O_2 ¹⁴, respectively. This residue is located in a loop characteristic of the AAO family (residues 395–406 in *P. eryngii* AAO), which partially covers the access to the active site²⁰ (Fig. 1). Phe397 has been proposed to act as a barrier that prevents the free diffusion of molecules —substrates and/or products— in and out of the catalytic pocket protecting the flavin environment of AAO. Previous computational studies of substrate migrations into the active site —using the PELE algorithm²¹— suggested that Phe397 swung along with the alcohol substrate, helping it to reach the catalytic pocket¹⁹ (www.youtube.com/watch?v=CqSDn5OmagI). These results have encouraged the investigation of the role of Phe397 in AAO catalysis.

In this work, kinetic, ligand binding and diffusion studies on AAO and several mutated variants have been performed to enlighten our understanding on the role of Phe397 in catalysis. Moreover, molecular dynamics and new ligand diffusion computational studies using adaptive-PELE —a new version of the PELE algorithm that avoids the metastability of the ligands, thus saving number of processors and computation time²²— have been employed to shed light on the catalytic implications of substituting Phe397 in AAO. Experimental findings and computational results agree very nicely and prove the applicability of the new adaptive-PELE to modelling the dynamic nature of protein-ligand interactions and its ability to unveil the mechanisms underlying binding in complex systems.

Results

Spectral properties and steady-state kinetics of AAO and its Phe397 variants. Wild-type recombinant AAO (hereinafter native AAO) and its F397A, F397L, F397Y and F397W variants were purified as holoproteins, after *Escherichia coli* expression and *in vitro* activation. The UV-visible spectra of all variants showed the typical flavin bands around 460 and 385 nm and shoulder around 500 nm (data not shown), indicating proper folding around the cofactor. 280 nm/460 nm absorbance ratios around 10–11 showed that the FAD cofactor was in the oxidized state and correctly incorporated into the protein for all variants.

The redox state of the cofactor during turnover was investigated by following the spectra of the AAO variants during oxidation of *p*-methoxybenzyl alcohol (initially saturated with air) until the enzyme was completely reduced. In the cases of native AAO and the majority of the variants —F397W, F397A and F397L— the most abundant species during turnover (initial reaction phase) is the oxidized enzyme, whereas the reduced species is the predominant for F397Y (Fig. 2).

Bi-substrate steady-state kinetics —measured as *p*-anisaldehyde production from *p*-methoxybenzyl alcohol at different O_2 concentrations— revealed remarkable differences among native AAO and its F397 variants (Table 1). First, in all four variants, contrary to the native AAO¹⁶, the kinetics best fitted equation (1) describing a ping-pong mechanism, as revealed by the Hanes-Woolf plots of its bi-substrate kinetics (Supplementary Fig. S1). Then, all variants, with the exception of F397L, showed 2–3 fold lower turnover rates (k_{cat} values) than the native enzyme. Regarding affinity for the alcohol substrate, the F397Y variant showed the same Michaelis constant ($K_{m(A)}$) as

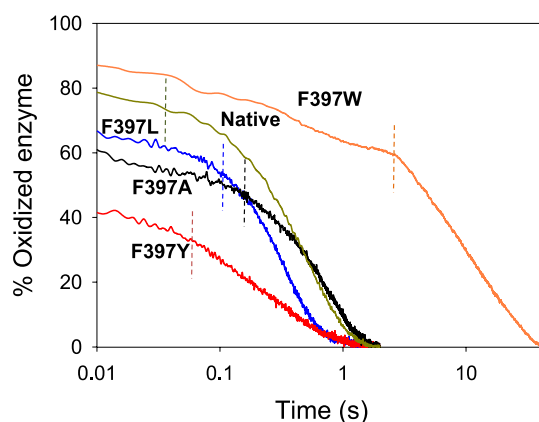


Figure 2. Redox state during turnover of native AAO and its Phe397 variants. Native (green), F397W (orange), F397L (blue), F397A (black) and F397Y (red) Native protein and its variants ($\sim 10 \mu\text{M}$) were mixed with an excess of *p*-methoxybenzyl alcohol in 50 mM sodium phosphate pH 6.0 at 25 °C under aerobic conditions. Lines show the time course of absorbance changes at the maxima of the flavin band I (in the 459 and 463 nm range, depending on the variant). Dashed lines separate the turnover phase from the drop in absorbance due to the depletion of O_2 .

	k_{cat} (s^{-1})	$K_{\text{m(Al)}}$ (μM)	$k_{\text{cat}}/K_{\text{m(Al)}}$ ($\text{s}^{-1}\cdot\text{mM}^{-1}$)	$K_{\text{m(Ox)}}$ (μM)	$k_{\text{cat}}/K_{\text{m(Ox)}}$ ($\text{s}^{-1}\cdot\text{mM}^{-1}$)
AAO ¹	129 ± 5	25 ± 3	5160 ± 650	348 ± 36	371 ± 41
F397Y	48 ± 1	25 ± 1	1920 ± 65	94 ± 4	512 ± 20
F397W	68 ± 1	280 ± 8	240 ± 5	90 ± 4	718 ± 10
F397A	66 ± 1	54 ± 1	1224 ± 32	500 ± 10	133 ± 3
F397L	115 ± 1	226 ± 4	506 ± 10	190 ± 4	610 ± 13

Table 1. Steady-state kinetic constants of native AAO and its Phe397 variants. The constants for *p*-methoxybenzyl alcohol (Al) and O_2 (Ox) were measured as the *p*-anisaldehyde produced in bi-substrate kinetic experiments, performed in 50 mM sodium phosphate (pH 6.0) at 12 °C. ¹From Ferreira *et al.*¹³. Means and standard deviations estimated from the fit to equation (1). All kinetics were measured by triplicates.

the native protein, while the F397A, F397L and F397W substitutions increased $K_{\text{m(Al)}}$ by ~ 2 -, 9- and 11-fold, respectively. Consequently, the F397Y, F397A, F397L and F397W variants were ~ 3 -, 4-, 10- and 21-fold less efficient ($k_{\text{cat}}/K_{\text{m}}$) oxidizing the alcohol substrate compared to the native AAO. Regarding the affinity for O_2 , all the variants, with the exception of F397A, showed increased affinity (lower $K_{\text{m(Ox)}}$) with regard to the native protein. Thus, F397Y, F397L and F397W were 1.4-, 1.6-, and 2-fold more efficient using O_2 as an electron acceptor than the native protein. Finally, steady-state constants calculated from anisaldehyde release (as in previous experiments) were compared with those obtained from H_2O_2 release (in additional kinetic analyses under atmospheric O_2 saturation). In all variants, k_{cat} and $K_{\text{m(Al)}}$ tended to be almost identical using both approaches (Supplementary Table S1), as it had been previously reported for native AAO²³.

Rapid kinetics of the two half-reactions for the Phe397 variants. In the light of the above results, the reductive and oxidative half-reactions of the F397 variants were analyzed to unveil the rate-limiting step during catalysis. The spectra collected during the reductive half-reactions of all the variants indicated an essentially irreversible two-electron reduction of the flavin, in agreement with the previously reported hydride transfer reaction for the native AAO¹⁶. Global analyses of the spectral evolution were fitted to a one-step model ($\text{A} \rightarrow \text{B}$) in all cases (Fig. 3). The values of the observed rate constants (k_{obs}) at different substrate concentrations exhibited a hyperbolic dependence on the alcohol concentration (Supplementary Fig. S2) that allowed the determination of the reduction rate constant (k_{red}) and the dissociation constant ($K_{\text{d(Al)}}$) upon fitting to equation (3). For native AAO (spectral changes not shown in Fig. 3) and the F397A and F397L variants, the k_{red} values (Table 2) were of the same range of the previously determined k_{cat} values (Table 1) indicating that the reductive half-reaction is the rate-limiting step in catalysis. Nevertheless, variants F397Y and F397W showed k_{red} values 3- and 2-fold higher than the respective turnover rates, suggesting that reductive half-reaction is not the limiting step in these variants. For all the F397 variants, $K_{\text{d(Al)}}$ values were similar to the $K_{\text{m(Al)}}$ estimated under steady-state conditions.

The oxidative half-reactions of the four Phe397 variants, and native AAO (not shown), fitted two-step model equations ($\text{A} \rightarrow \text{B} \rightarrow \text{C}$) describing a biphasic pattern (Fig. 4) where the first phase accounts for more than

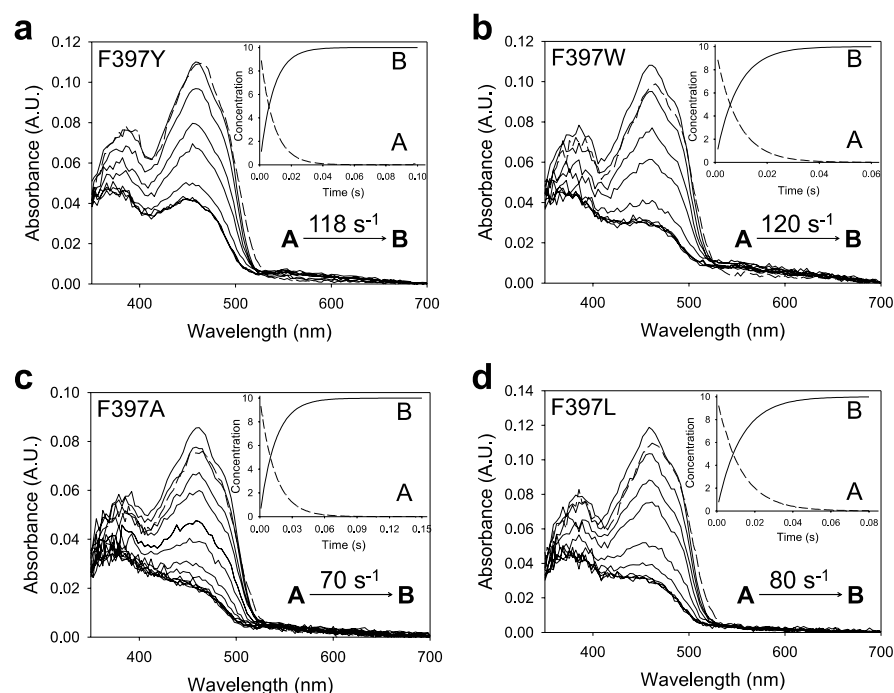


Figure 3. Time course of reduction of Phe397 variants with *p*-methoxybenzyl alcohol. (a) Spectra of F397Y mixed with 600 μM substrate measured at 0.003, 0.005, 0.01, 0.015, 0.02, 0.025, 0.06 and 0.1 s after mixing. (b) Spectra of F397W mixed with 2400 μM substrate recorded at 0.003, 0.01, 0.02, 0.03, 0.05, 0.07, 0.2 and 0.5 s. (c) F397A reduction spectra with 1200 μM substrate at 0.003, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.08 and 0.3 s. (d) Spectra of F397L mixed with 2400 μM substrate at 0.003, 0.01, 0.02, 0.03, 0.05, 0.07, 0.1, 0.13 and 0.4 s. Enzyme ($\sim 10 \mu\text{M}$) reactions were performed in 50 mM sodium phosphate, pH 6.0, at 12 $^{\circ}\text{C}$. Dashed lines correspond to the oxidized enzymes before mixing. Data were globally fitted to a single-step model described from initial species A to final species B (shown in insets). Estimated k_{obs} is represented in each panel.

	k_{red} (s^{-1})	$K_{\text{d(A)}}$ (μM)	$k_{\text{red}}/K_{\text{d}}$ ($\text{s}^{-1}\cdot\text{mM}^{-1}$)	$^{\text{app}}k_{\text{ox}}$ ($\text{s}^{-1}\cdot\text{mM}^{-1}$)
AAO	115 ± 3	31 ± 2	3710 ± 258	770 ± 40
F397Y	150 ± 3	41 ± 3	3660 ± 277	770 ± 70
F397W ¹	124 ± 3	292 ± 17	425 ± 10	689 ± 92
F397A	69 ± 1	61 ± 2	1130 ± 40	78 ± 4
F397L	87 ± 1	180 ± 7	483 ± 20	340 ± 10

Table 2. Transient-state kinetic constants for the reductive and oxidative half-reactions of AAO and its Phe397 variants. The constants were measured using stopped-flow rapid spectrophotometry in 50 mM sodium phosphate (pH 6.0) at 12 $^{\circ}\text{C}$ under anaerobic conditions. ¹The F397W constants for the first phase of the oxidative half-reaction show a hyperbolic dependence on O_2 concentration (in contrast to the other variants) with $k_{\text{ox}}/K_{\text{d(ox)}}$ and k_{ox} values of $689 \pm 92 \text{ mM}^{-1}\text{s}^{-1}$ and $156 \pm 12 \text{ s}^{-1}$ respectively, estimated from fit to equation (5). Means and standard deviations estimated from the fits to equations (3), (4) and (5). All kinetics were measured by triplicates.

70–80% of the total amplitude. For the F397Y, F397L and F397A variants, this $k_{\text{obsA} \rightarrow \text{B}}$ was linearly dependent on O_2 concentration, allowing the determination of a second-order rate constant ($^{\text{app}}k_{\text{ox}}$) that was similar, 2- or 10-fold slower than that of native AAO, respectively (Table 2). For the F397L and F397A variants, a reverse rate constant was observed, $k_{\text{rev}} \sim 15 \text{ s}^{-1}$, which corresponds to the intercept of the y-axis (Supplementary Fig. S3A). On the contrary, the $k_{\text{obsA} \rightarrow \text{B}}$ for F397W showed a hyperbolic dependence on O_2 followed by flavin reoxidation (Supplementary Fig. S3B). The estimated transient-state second order rate constant, $k_{\text{ox}}/K_{\text{d(ox)}}$, $\sim 689 \text{ s}^{-1}\text{mM}^{-1}$ (Table 2 footnote) agreed with the catalytic efficiency determined under steady-state conditions (Table 1).

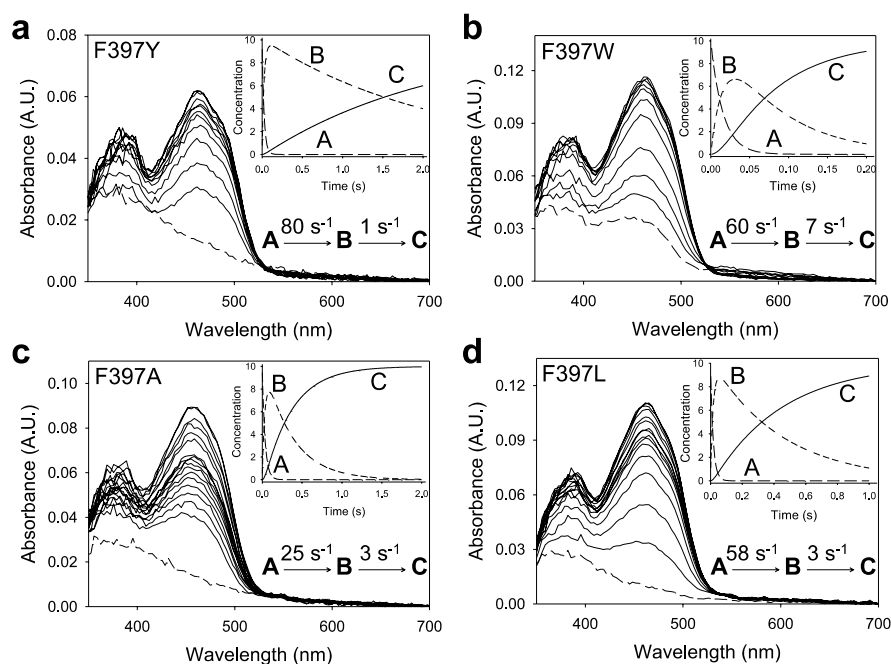


Figure 4. Time course of reoxidation of Phe397 variants with O_2 . (a) F397Y spectra measured at 0.002, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.5, 1 and 1.5 s after mixing. (b) F397W recorded at 0.001, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.18 and 0.2 s. (c) F397A recorded at 0.002, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.5, 1, and 1.5 s. (d) F397L recorded at 0.002, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.3, 0.5, 0.7, 0.9 and 1 s. Enzyme ($\sim 10 \mu\text{M}$) reactions with O_2 ($136 \mu\text{M}$) were performed in 50 mM sodium phosphate, pH 6.0, at 12°C . Dashed lines correspond to the reduced enzymes. Insets show the evolution of species A, B and C after data fitting to a two-step process. The estimated k_{obs} for each phase is represented in each panel.

The second phase for all variants was too slow to be relevant for catalysis ($k_{\text{obs}2} < 7 \text{ s}^{-1}$), although it was independent of O_2 concentration for the F397Y and F397W variants and dependent on it in the case of the F397L and F397A variants (Supplementary Fig. S3A, inset). The O_2 -independent slow phase in native AAO reoxidation had been previously attributed to the presence of damaged protein ensuing the stopped-flow experiments²⁴.

Studies on AAO:*p*-anisic acid complex formation and dissociation. The differences between the k_{cat} and k_{red} values for both F397Y and F397W encouraged us to investigate whether product release has an effect on turnover. The study could not be carried out with *p*-anisaldehyde because the formation of the enzyme-aldehyde complex is too rapid, which prevents its detection by the stopped flow equipment. Therefore, this study (Fig. 5) was performed with the final product, *p*-anisic acid, whose crystallographic complex with AAO was recently solved (PDB 5OC1)¹⁷. The rate constants for complex formation (k_{for}) and dissociation (k_{dis}) of native AAO and the F397Y and F397W variants are shown in Table 3. The k_{for} for F397Y and F397W were 2-fold and 2 orders of magnitude lower than that of the native enzyme, respectively. Differences in k_{dis} were also observed, since the k_{dis} values for these variants were at least 10-fold slower than that of the native AAO. The fast diffusion of the ligand in and out the active site of the F397L and F397A variants most likely impeded estimation of these parameters. The $K_{\text{d(Ac)}}$ values calculated from the corresponding $k_{\text{dis}}/k_{\text{for}}$ ratios for native AAO and the F397Y and F397W variants (150, 31 and $271 \mu\text{M}$ respectively) agree well with the values obtained by differential spectrophotometry (Supplementary Fig. S4 and Table 3). F397Y showed the smallest $K_{\text{d(Ac)}}$, which means that it is the variant that more tightly binds *p*-anisic acid. On the contrary, the $K_{\text{d(Ac)}}$ for the F397L and F397A variants could not be estimated by differential spectrophotometry since saturation could not be attained in agreement with the above results.

AAO molecular dynamics and ligand diffusion simulations. MD simulations were carried out to study the dynamical behaviour of the five protein systems (native enzyme and its Phe397 variants). Loops are the secondary structures that most likely adopt different conformational structures. Hence, the motions of the loops Gln395-Thr406 and Ser89-Met95, where two of the main gate residues —Phe397 and Tyr92— are located, were investigated. However, the root-mean-square-deviation values obtained, as a function of time, were low (data not shown) indicating that there are only moderate displacements of these secondary structures along the dynamics of the protein.

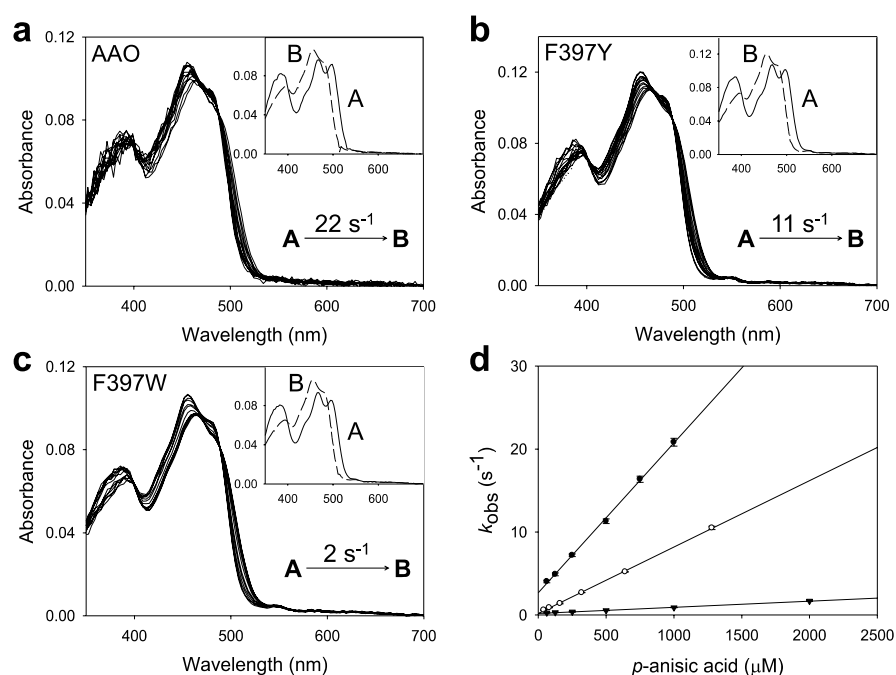


Figure 5. Spectral changes during formation of AAO:*p*-anisic acid complex. (a) Native AAO, mixed with 0.5 mM of the ligand at 0.003, 0.01, 0.02, 0.03, 0.04, 0.05, 0.065, 0.085, 0.1, 0.13, 0.15, 0.2 and 0.4 s. (b) F397Y, mixed with 0.65 mM of *p*-anisic acid measured after 0.005, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3 and 4 s, and (c) F397W mixed with *p*-anisic acid 1 mM recorded at 0.005, 0.02, 0.04, 0.05, 0.07, 0.1, 0.2, 0.4, 0.5, 0.7, 1, 2 and 5 s. Insets show disappearance of initial species A and formation of final species B. The rate constant obtained from the global fitting of the spectral changes is shown. (d) Fits of the obtained k_{obs} for each variant and ligand concentration to equation (6). Native AAO (filled circles), and F397Y (open circles) and F397W (inverted triangles) variants.

	$K_{d(Ac)}$ (μM)	k_{for} ($\text{s}^{-1}\cdot\text{mM}^{-1}$)	k_{dis} (s^{-1})
AAO	170 ± 5	18.0 ± 0.3	2.7 ± 0.17
F397Y	25 ± 2	7.7 ± 0.1	0.24 ± 0.04
F397W	200 ± 12	0.70 ± 0.01	0.19 ± 0.01
F397A	nd	nd	nd
F397L	nd	nd	nd

Table 3. Dissociation constants, and formation/dissociation rates of *p*-anisic acid complexes. The dissociation constants ($K_{d(Ac)}$) of *p*-anisic acid complexes with AAO and its Phe397 variants were measured by differential spectroscopy, while the complex formation (k_{for}) and dissociation (k_{dis}) rates were measured by transient-state kinetics spectroscopy. All experiments were performed in 50 mM sodium phosphate (pH 6.0) at 12 °C. nd, not determined (too fast complex formation and dissociation). Means and standard deviations from the fit to equations (7) or (6). All data were measured by triplicates.

Despite the small loop movements, a closer look at the distances between the gate residues indicates that the F397Y and F397W substitutions induce significantly shorter values (Supplementary Fig. S5A). Moreover, inspection of the protein structures shows that the side chains of the F397Y and F397W variants establish hydrogen bonds with Tyr92 with a frequency of 20% and 5%, respectively (Supplementary Fig. S5B), while the aliphatic variants and the native enzyme cannot develop such hydrogen bonds resulting in larger distances between gate residues. Altogether, the MD data suggests local constraints induced by the tyrosine and tryptophan substitutions in the gate region.

The diffusion of products —*p*-anisaldehyde and *p*-anisic acid— and reactants —*p*-methoxybenzyl alcohol and O_2 — was also investigated with adaptive-PELE to explain the changes in kinetics observed for the mutants. The trajectory analysis of the acidic product in the aromatic variants clearly indicates a stabilization of the local

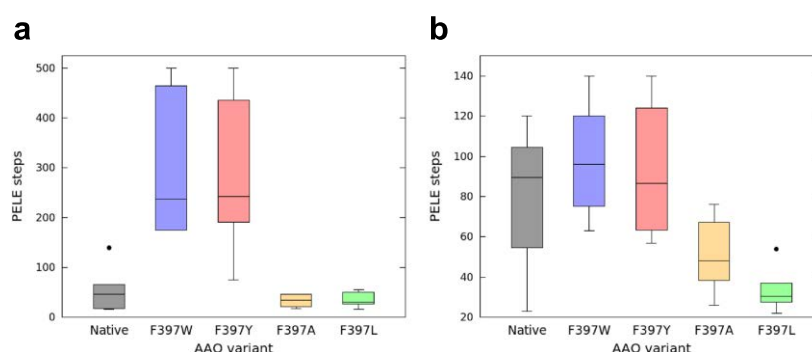


Figure 6. Boxplots of PELE steps required for ligand release from AAO and its Phe397 variants. **(a)** *p*-Anisic acid. **(b)** *p*-Anisaldehyde. Number of PELE steps is directly related to time. Grey, native AAO; blue, F397W; red, F397Y; yellow, F397A; and light green, F397L variants. Boxes contain 50% of the results (two quartiles), horizontal lines inside them indicate the mean value; while the upper and lower whiskers (vertical lines) contain the remaining quartiles (50%). The overall height of boxes (whiskers included) is indicative of the spread of the results. Isolated points represent results significantly different from the rest of data values.

minima at the gate, before ligand release, which does not take place during release of the aldehyde products. This energy minimum involves interaction between the carboxylic group and three amino acids, Ser393, Gln395 and Ser411, placed inside the catalytic cavity. The aromatic variants required in average ~3-fold and ~7-fold more simulation steps to release the aldehyde and acid, respectively, than the aliphatic ones (Fig. 6). Moreover, the number of PELE steps required to release both products also tends to be higher in the case of the aromatic mutants than for the native enzyme, especially for anisic acid due to the interactions of the carboxylic group with the three residues mentioned above.

Furthermore, the distances among any of the atoms of the O_2 substrate and both the C4a locus of reduced FAD and the N_{ϵ_2} of His502 (acting as a catalytic base in the reductive half reaction) were calculated for all the variants. Results suggest that the catalytic distances are similar in the native enzyme and variants, except for F397W that shows higher population at favourable—that is, shorter—distances (Fig. 7).

Finally, the energetic profiles of the *p*-methoxybenzyl alcohol catalytic distances (hydroxyl-to-His502 and proR-H-to-flavin) in all protein systems were studied, as a measure of the binding effectiveness of the alcohol in the active centre. The energy landscape shows an additional energetic favourable region only for the F397W. In this new minimum, the side chain of the tryptophan interacts via hydrogen bond to the hydroxyl group of the alcohol substrate, resulting in an inefficient catalytic pose (Supplementary Fig. S6). This is in agreement with the experimental values, where the $K_{m(Al)}$ and $K_{d(Al)}$ are almost 11-fold higher for the F397W variant compared to the native protein.

Discussion

Phe397 is situated at a loop characteristic of the AAO family, in a region prone to harbouring insertions and deletions amongst the members of the GMC superfamily. Representatives of this superfamily are monomeric or multimeric proteins, with the entrance to their active site partially covered by the adjacent monomer in the second case. Given their monomeric nature, AAOs and a few other GMCs have developed loop structures to control diffusion of molecules into their active sites. Among them, choline oxidase²⁵, cholesterol oxidase²⁶ and cellobiose dehydrogenase²⁷ are notable examples, whose insertions are even longer than the AAO insertion. The latter insertion partially forms the 395–406 loop in *P. eryngii* AAO and encloses the active-site cavity from the outer environment (Fig. 1)¹⁸. Previous computational studies hinted that Phe397 oscillates with the substrate as a mechanism of gating¹⁹, similarly to what has been reported for phenylalanine residues in P450 enzymes²⁸. A homologous phenylalanine residue is conserved in 50% of all the putative AAO sequences from basidiomycete genomes available at JGI (<https://jgi.doe.gov>). The multiple roles of this phenylalanine residue in catalysis, as revealed by studies on the *P. eryngii* model AAO, are discussed below.

Phe397 favours the correct positioning of the alcohol and its oxidation at the active site, as evidenced by the steep decrease of catalytic efficiency in the mutated variants. Such effect is most noticeable in F397W, which shows the lowest $k_{cat}/K_{m(Al)}$ value due to: i) the constrained distance between gate residues, and ii) the formation of a H bond with its reactive hydroxyl group that disrupts the substrate's catalytic poses. The introduction of a bulkier residue at the Phe501 and Tyr92 positions of AAO also produced similar steric hindrances for catalysis^{13,14,17}. Remarkably, the drop of the catalytic efficiency in F397Y is due to its small turnover rates, as affinity for alcohol ($K_{m(Al)}$ and $K_{d(Al)}$) remains invariable in this case. Furthermore, the reduction rates being higher than turnover in F397Y and F397W, indicates that, unlike native enzyme, the reductive half-reaction is not the rate-determining step in these two variants, as discussed below.

Phe397 also plays a role in flavin reoxidation by either compressing the active site to reduce the free diffusion of O_2 , or altering its redox environment. Evidence comes from the low oxidation rates estimated for the aliphatic mutants, in contrast to the ones obtained for the aromatic variants and the native AAO, which possess bulkier

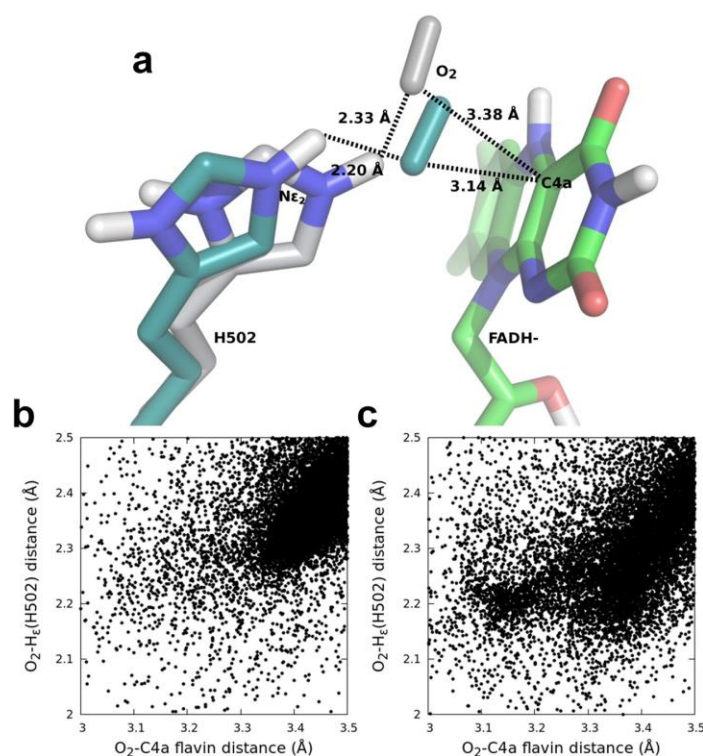


Figure 7. O_2 diffusion in native AAO and its F397W variant. (a) Native and F397W final catalytic poses (CPK-colored molecules with grey and cyan carbon atoms, respectively) showing catalytic distances (in Å) to FAD C4a and His502. (b and c) Frequency of O_2 poses as a function of the distances from O_2 to His502 and C4a of native and the F397W variant, respectively.

residues that constrain the catalytic pocket. However, the oxidative half-reaction does not limit the catalytic cycle in the aliphatic variants, as k_{cat} and k_{red} values are similar and the constants measured as H_2O_2 release are identical as those measured by aldehyde formation. Alteration of the FAD environment and electrochemistry might be involved in the decreased efficiency for oxidation of the F397A cofactor. Regarding the F397Y and F397W variants, they show high $^{app}k_{ox}$, similar to the native protein. In the case of F397W variant, the obtained k_{ox} and k_{red} are similar indicating that the reductive and oxidative half-reactions are almost balanced, which agrees with its redox state during turnover. Computational simulations of F397W reoxidation revealed that withdrawal of His502 due to the presence of the tryptophan increases the ability of this variant for properly positioning the oxygen molecule during flavin reoxidation, bringing O_2 and C4a closer than in the native enzyme (Fig. 7). However, the introduction of this bulkier residue could be hindering the oxygen diffusion into the active-site¹⁹ and thus causing the unusual saturation behaviour observed for this variant with increasing oxygen concentration. This saturation profile has been observed in cholesterol oxidase from *Brevibacterium sterolicum* related to the interconversion between open and closed channel enzyme conformations. This event regulated the oxygen diffusion and constituted the rate-limiting step preceding flavin reoxidation^{29,30}. Regarding F397A and F397L, they display slower reoxidation rates probably due to the absence of an aromatic residue in that position. Moreover, both variants display slow phases that are dependent on O_2 concentration, that could reveal two different kinetic processes, the H and H^+ transfers recently reported to take place in separate kinetic steps in native AAO²⁴.

Finally, Phe397 also plays a role in controlling the product release from the active site, with this catalytic step limiting the turnover of the F397Y and F397W variants, while the F397A and F397L variants (and native AAO) let the substrate and product diffuse easily in and out of the active site. Furthermore, ligand migration studies with both *p*-anisic acid and *p*-anisaldehyde indicate that product release is delayed due to the reduced gated space—which forces the aromatic side chains to be displaced—and the hydrogen bond between the tyrosine or tryptophan and Tyr92 in the aromatic variants. The native enzyme also requires longer times for product release because its side chain must move to let it out, although there is no hydrogen bonding between the Phe397 and Tyr92. This is in agreement with the 10-fold lower k_{dis} for the aromatic variants than for native AAO. Flavoenzymes with catalytic cycles limited or partially limited by the product release have been reported, such as D-amino acid oxidase³¹, cyclohexanone monooxygenase³² or nitroalkane oxidase³³. The case of amadoriase I is also relevant, since

its catalysis is partially limited by the product release, although the oxidation of the enzyme takes place through an ordered ternary complex of enzyme, product and O₂³⁴.

In conclusion, Phe397 plays a central role in the catalysis of *P. eryngii* AAO, which is reinforced by its high conservation among other sequences of characterised and putative AAOs from fungal genomes. As drawn from the experimental and computational results, this residue supports both half-reactions of the enzyme. During the reductive reaction, it helps the substrate attain a catalytically relevant position and facilitates product release from the active site. Analysis of the oxidative reaction suggests that the presence of an aromatic residue at this position is important to modulate the flavin environment and to reduce the space in the active site enabling the enzyme to reduce O₂ to H₂O₂. Experimental and computational results complement each other well and, thus, highlight the suitability of the recently developed adaptive-PELE²² for the study of protein binding to different ligand types.

Methods

Reagents. Glucose oxidase type VII from *Aspergillus niger*, glucose, *p*-methoxybenzyl alcohol and *p*-anisic acid were purchased from Sigma-Aldrich.

Enzyme expression and purification. Native AAO from *P. eryngii* (GenBank accession number **AF064069**) and its mutated variants F397A, F397Y, F397W and F397L were heterologously expressed in *E. coli* W3110 as recombinant proteins using the pFLAG1 vector. The above-mentioned AAO variants were prepared using the QuickChange[®] site-directed mutagenesis kit.

PCR reactions were run with the following oligonucleotides harbouring the desired mutations (underlined nucleotide/s in bold triplets indicated below) and the AAO gene inserted in the vector as a template: (i) F397Y: 5'-CTTTTCCAACCAATGGT**ACC**ACCCAGCTATCCCTCG-3', (ii) F397A: 5'-CTTTTCCAACCAATGG**GGC**CCACCCA GCTATCCCTCG-3', (iii) F397W: 5'-CTTTTCCAACCAATGGT**GGC**ACCCAGCTATCCCTCG-3', and (iv) F397L: 5'-CTTTTCCAACCAATGGT**TGC**ACCCAGCTATCCCTCGC3', along with their reverse complementary counterparts. The AAO variants were expressed as inclusion bodies, subsequently *in vitro* activated in the presence of FAD and purified by anion-exchange chromatography as previously described³⁵.

Molar absorbance coefficients of the variants were calculated by heat denaturation of the proteins and determination of the FAD released. The coefficients for native AAO and the F397Y, F397A, F397W and F397L variants are $\epsilon_{463\text{ nm}} = 11050\text{ M}^{-1}\text{ cm}^{-1}$, $\epsilon_{462\text{ nm}} = 10415\text{ M}^{-1}\text{ cm}^{-1}$, $\epsilon_{459\text{ nm}} = 9432\text{ M}^{-1}\text{ cm}^{-1}$, $\epsilon_{462\text{ nm}} = 9952\text{ M}^{-1}\text{ cm}^{-1}$ (at 462 nm), and $\epsilon_{462\text{ nm}} = 12492\text{ M}^{-1}\text{ cm}^{-1}$, respectively.

Steady-state kinetics. Bi-substrate kinetics of the four AAO variants were measured recording *p*-anisaldehyde production ($\Delta\epsilon_{285} = 16950\text{ M}^{-1}\text{ cm}^{-1}$), varying both the concentrations of *p*-methoxybenzyl alcohol and O₂ in a spectrophotometer. Reactions were performed in screw-cap cuvettes, where different O₂ concentrations in 0.1 M phosphate at pH 6.0, were attained by bubbling several O₂/N₂ mixtures inside a thermostatic bath at 12 °C. Reactions were triggered by the addition of the alcohol substrate (5–500 μM) and the enzyme (3–5 nM) using microsyringes. The spectrophotometer was maintained at the desired temperature owing to the use of a thermostatic bath.

Initial rates were obtained from the linear phase of the aldehyde production and were calculated as the change in absorbance over time. Kinetic constants were estimated by fitting the k_{obs} to equation (1) describing a ping-pong mechanism:

$$\frac{\nu}{e} = \frac{k_{\text{cat}}AB}{K_{m(\text{Al})}B + K_{m(\text{Ox})}A + AB} \quad (1)$$

where ν stands for the initial velocity, e is the enzyme concentration, k_{cat} is the catalytic constant, A stands for the alcohol concentration, B represents O₂ concentration and $K_{m(\text{Al})}$ and $K_{m(\text{ox})}$ are the Michaelis constants for *p*-methoxybenzyl alcohol and O₂, respectively.

Kinetic constants estimated as *p*-anisaldehyde and H₂O₂ release were compared under atmospheric O₂ saturation conditions, at 25 °C. *p*-Anisaldehyde was measured as detailed above, while H₂O₂ release was estimated by coupling the reaction of a horseradish peroxidase (6 U·mL⁻¹) and AmplexRed[®] (60 nM), which uses the H₂O₂ produced by AAO to give coloured resorufin ($\Delta\epsilon_{563} = 52000\text{ M}^{-1}\text{ cm}^{-1}$). Catalytic constants were estimated by fitting to a Michaelis-Menten equation for one substrate in both cases (equation (2)):

$$\frac{\nu}{e} = \frac{k_{\text{cat}}A}{K_{m(\text{Al})} + A} \quad (2)$$

Stopped-flow measurements. Experiments were performed using a stopped-flow spectrophotometer from Applied Photophysics Ltd. model SX17.MV.

Enzyme-monitored turnover experiments were carried out by mixing AAO (~10 μM) with exceedingly saturating *p*-methoxybenzyl alcohol concentrations (at least 10-fold the K_m for each of the variants), under air-saturated conditions, at 25 °C. Spectral evolution of the enzymes during redox turnover was recorded between 350 and 700 nm with a photodiode array (PDA) detector.

Studies on the reductive half-reaction were performed upon mixing AAO with increasing concentrations of *p*-methoxybenzyl alcohol (0.018–0.6 mM) under anaerobic conditions. The stopped-flow apparatus was made anaerobic by flushing sodium dithionite through the system, which was then rinsed out with O₂-free buffer. All buffers, substrates and the enzyme were poured into glass tonometers that were subsequently subjected to 20–25 cycles

of evacuation and argon (Ar) flushing in order to remove O₂. To ensure anaerobiosis, glucose (10 mM) and glucose oxidase (10 U·mL⁻¹) were added after some vacuum-Ar cycles. Measurements were recorded using the PDA detector, in 50 mM sodium phosphate (pH 6.0) at 12 °C. Observed rate constants (k_{obs}) were calculated by global fitting of the spectra with Pro-K software to a single exponential equation. Those averaged k_{obs} at each substrate concentration were then non-linearly fitted to equation (3), describing hyperbolic substrate dependence of k_{obs} :

$$k_{\text{obs}} = \frac{k_{\text{red}}A}{K_{\text{d(A)}} + A} + k_{\text{rev}} \quad (3)$$

where k_{red} and k_{rev} represent the reduction rate constant at infinite substrate concentration and its reverse reaction, respectively; A stands for the alcohol concentration; and K_{d} is the dissociation constant.

The oxidative half-reaction was investigated by mixing reduced AAO with increasing O₂ concentrations. Procedures were as explained above for the reductive half-reaction, except that, in this case, AAO and glucose were put into a tonometer bearing a side-arm, where *p*-methoxybenzyl alcohol (1.3-fold the concentration of AAO) was poured along with glucose oxidase. After the required vacuum-Ar cycles, enzyme and substrate were mixed before being mounted onto the stopped-flow equipment. Reactions were measured with both the PDA and the monochromator detectors at 12 °C. k_{obs} were obtained by either global fitting of the spectra or fitting the monochromator traces to exponential equations describing two-step and three-step processes. Fitting averaged k_{obs} either to equation (4) that describes a linear dependence on O₂ concentration or equation (5) that describes hyperbolic saturation with increasing O₂ concentration allowed the estimation of the apparent second-order rate constant for reoxidation ($^{\text{app}}k_{\text{ox}}$) and the first-order rate (k_{ox}) and second-order constants for reoxidation ($k_{\text{ox}}/K_{\text{d(ox)}}$), respectively:

$$k_{\text{obs}} = ^{\text{app}}k_{\text{ox}}\text{O}_2 + k_{\text{rev}} \quad (4)$$

$$k_{\text{obs}} = \frac{k_{\text{ox}}/K_{\text{d(ox)}}\text{O}_2}{1 + (k_{\text{ox}}/K_{\text{d(ox)}}\text{O}_2)/k_{\text{ox}}} \quad (5)$$

Estimation of the rates of the AAO:*p*-anisic acid complex formation and dissociation were performed by analyzing spectral changes upon mixing enzyme (~10 μM) with different concentrations of the ligand (0.04–2 mM) at 12 °C. Data were globally fitted to an equation describing a one-step process. The obtained k_{obs} were linearly depended on the ligand concentration and were, thus, fitted to equation (6):

$$k_{\text{obs}} = k_{\text{for}}L + k_{\text{dis}} \quad (6)$$

in which k_{for} stands for the second-order rate constant for the complex formation; L represents the ligand concentration, and k_{dis} is the rate constant for the complex dissociation.

Spectral characterization of the AAO-*p*-anisic acid complex. The affinity of the native AAO and its variants for *p*-anisic acid was assessed by titrating the enzyme with increasing concentrations of the ligand in 50 mM sodium phosphate (pH 6.0) at 12 °C. Spectral changes were recorded using a spectrophotometer and their magnitude upon complex formation was fitted to equation (7), which accounts for a 1:1 stoichiometry, as a function of *p*-anisic acid concentration:

$$\Delta A = \frac{\Delta\epsilon(E + L + K_{\text{d}}) - \Delta\epsilon\sqrt{(E + L + K_{\text{d}})^2 - 4EL}}{2} \quad (7)$$

in which ΔA accounts for the observed change in absorbance; $\Delta\epsilon$ represents the maximal absorption difference in each of the spectra; K_{d} is the dissociation constant; and E and L , the enzyme and *p*-anisic acid concentrations.

Molecular dynamics (MD) studies. The protonation states of the titratable amino acids present at the AAO crystal structure (PDB 3FIM) were set up using PROPKA³⁶ of *Protein Preparation Wizard*³⁷ and checked with the H++ server³⁸. The selected mutations (F397W, F397Y, F397A and F397L) were introduced and all 3D structures were minimized with Prime^{39,40}. The overall systems were solvated and neutralized using Desmond⁴¹ with the SPC water model and 0.15 M of NaCl.

50 ns of MD simulation were performed with Desmond⁴¹ for each protein structure using the default Desmond's protocol. The temperature was regulated with the *Nosé-Hoover* chain thermostat^{42–44} with a relaxation time of 1 ps. The pressure was controlled with the *Martyna-Tobias-Klein barostat*⁴⁵ with isotropic coupling and a relaxation time of 2 ps. RESPA integrator^{45,46} was used with bonded, near and far time steps of 2.0, 2.0 and 6.0 ps, respectively. Furthermore, a 9 Å cut-off was employed for nonbonded interactions with the smooth particle *mesh Ewald method*⁴⁷.

Ligand diffusion. The minimized models described above were also used to study the ligand diffusion. The acid product was manually docked inside the cavity, next to Phe397 and Tyr92, while reactants were docked in their catalytic position. The FAD cofactor was protonated to its semiquinone state for the O₂ simulations, while it remained in its quinone state for the acid, aldehyde and alcohol diffusion studies.

The FAD cofactor was optimized with QM/MM at the *M06/6-31G** level of theory using Qsite⁴⁸ and the atomic charges were obtained from the electrostatic potential (ESP). Ligand geometries were optimized at the

same level of theory in gas phase and re-optimized using the PBF implicit solvent with the Jaguar software⁴⁹. The ligands were parameterized according to the OPLS 2005 force field, maintaining the ESP charges, and a rotamer library was build using *Macromodel*³⁷.

Ligand diffusion was modelled with the recent adaptive-PELE technique²², using an anisotropic network model (ANM) backbone perturbation⁵⁰ applied to all C α atoms, side chains re-sampling within 6 Å of the ligands, and a full energy minimization in each PELE step. The products were randomly translated and rotated within a spherical box of 22 Å of its initial position. The O₂ and the alcohol reactants were randomly roto-translated within a spherical box of 3 Å and 7 Å of its centre of mass initial position, respectively, to inquire about their poses in the active centre. Adaptive-PELE simulations involved ~50 epochs of 4 PELE steps. All simulations were performed using 32 processors. The O₂ simulations were executed during 72 h, while the other simulations were performed during 23 hours.

Data availability statement. All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

References

- Martínez, A. T. *et al.* Oxidoreductases on their way to industrial biotransformations. *Biotechnol. Adv.* **35**, 815–831 (2017).
- Hernández-Ortega, A., Ferreira, P. & Martínez, A. T. Fungal aryl-alcohol oxidase: A peroxide-producing flavoenzyme involved in lignin degradation. *Appl. Microbiol. Biotechnol.* **93**, 1395–1410 (2012).
- Ferreira, P. *et al.* Spectral and catalytic properties of aryl-alcohol oxidase, a fungal flavoenzyme acting on polyunsaturated alcohols. *Biochem. J.* **389**, 731–738 (2005).
- Guillén, F., Martínez, A. T. & Martínez, M. J. Substrate specificity and properties of the aryl-alcohol oxidase from the ligninolytic fungus *Pleurotus eryngii*. *Eur. J. Biochem.* **209**, 603–611 (1992).
- Carro, J., Serrano, A., Ferreira, P. & Martínez, A. T. Microbial Enzymes in Bioconversions of Biomass. Gupta, V. K. & Tuohy, M. G. (eds), pp. 301–322 (Springer, Berlin, Germany, 2016).
- Ruiz-Dueñas, F. J. & Martínez, A. T. Microbial degradation of lignin: How a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microbial Biotechnol.* **2**, 164–177 (2009).
- Guillén, F. & Evans, C. S. Anisaldehyde and veratraldehyde acting as redox cycling agents for H₂O₂ production by *Pleurotus eryngii*. *Appl. Environ. Microbiol.* **60**, 2811–2817 (1994).
- Gutiérrez, A., Caramelo, L., Prieto, A., Martínez, M. J. & Martínez, A. T. Anisaldehyde production and aryl-alcohol oxidase and dehydrogenase activities in ligninolytic fungi from the genus *Pleurotus*. *Appl. Environ. Microbiol.* **60**, 1783–1788 (1994).
- de Jong, E., Cazemier, A. E., Field, J. A. & de Bont, J. A. M. Physiological role of chlorinated aryl alcohols biosynthesized de novo by the white rot fungus *Bjerkandera* sp strain BOS55. *Appl. Environ. Microbiol.* **60**, 271–277 (1994).
- Carro, J. *et al.* 5-Hydroxymethylfurfural conversion by fungal aryl-alcohol oxidase and unspecific peroxygenase. *FEBS J.* **282**, 3218–3229 (2015).
- Carro, J. *et al.* Self-sustained enzymatic cascade for the production of 2,5-furandicarboxylic acid from 5-methoxymethylfurfural. *Biotechnol. Biofuels* **11**, 86 (2018).
- Hernández-Ortega, A. *et al.* Stereoselective hydride transfer by aryl-alcohol oxidase, a member of the GMC superfamily. *ChemBioChem* **13**, 427–435 (2012).
- Ferreira, P. *et al.* Aromatic stacking interactions govern catalysis in aryl-alcohol oxidase. *FEBS J.* **282**, 3091–3106 (2015).
- Hernández-Ortega, A. *et al.* Modulating O₂ reactivity in a fungal flavoenzyme: Involvement of aryl-alcohol oxidase Phe-501 contiguous to catalytic histidine. *J. Biol. Chem.* **286**, 41105–41114 (2011).
- Hernández-Ortega, A. *et al.* Role of active site histidines in the two half-reactions of the aryl-alcohol oxidase catalytic cycle. *Biochemistry* **51**, 6595–6608 (2012).
- Ferreira, P., Hernández-Ortega, A., Herguedas, B., Martínez, A. T. & Medina, M. Aryl-alcohol oxidase involved in lignin degradation: A mechanistic study based on steady and pre-steady state kinetics and primary and solvent isotope effects with two different alcohol substrates. *J. Biol. Chem.* **284**, 24840–24847 (2009).
- Carro, J., Martínez-Júlviz, M., Medina, M., Martínez, A. T. & Ferreira, P. Protein dynamics promote hydride tunnelling in substrate oxidation by aryl-alcohol oxidase. *Phys. Chem. Chem. Phys.* **19**, 28666–28675 (2017).
- Fernández, I. S. *et al.* Novel structural features in the GMC family of oxidoreductases revealed by the crystal structure of fungal aryl-alcohol oxidase. *Acta Crystallogr. D. Biol. Crystallogr.* **65**, 1196–1205 (2009).
- Hernández-Ortega, A. *et al.* Substrate diffusion and oxidation in GMC oxidoreductases: An experimental and computational study on fungal aryl-alcohol oxidase. *Biochem. J.* **436**, 341–350 (2011).
- Ferreira, P., Carro, J. & Serrano, A. & Martínez, A. T. A survey of genes encoding H₂O₂-producing GMC oxidoreductases in 10 Polyporales genomes. *Mycologia* **107**, 1105–1119 (2015).
- Borrelli, K. W., Vitalis, A., Alcantara, R. & Guallar, V. PELE: Protein energy landscape exploration. A novel Monte Carlo based technique. *J. Chem. Theory Comput.* **1**, 1304–1311 (2005).
- Lecina, D., Gilabert, J. F. & Guallar, V. Adaptive simulations, towards interactive protein-ligand modeling. *Sci. Rep.* **7**, 8466 (2017).
- Ferreira, P. *et al.* Kinetic and chemical characterization of aldehyde oxidation by fungal aryl-alcohol oxidase. *Biochem. J.* **425**, 585–593 (2010).
- Carro, J., Ferreira, P., Martínez, A. T. & Gadda, G. Stepwise hydrogen atom and proton transfers in dioxygen reduction by aryl-alcohol oxidase. *Biochemistry* **57**, 1790–1797 (2018).
- Xin, Y., Gadda, G. & Hamelberg, D. The cluster of hydrophobic residues controls the entrance to the active site of choline oxidase. *Biochemistry* **48**, 9599–9605 (2009).
- Vrielink, A. & Ghisla, S. Cholesterol oxidase: biochemistry and structural features. *FEBS J.* **276**, 6826–6843 (2009).
- Hallberg, B. M., Henriksson, G., Pettersson, G. & Divne, C. Crystal structure of the flavoprotein domain of the extracellular flavocytochrome cellobiose dehydrogenase. *J. Mol. Biol.* **315**, 421–434 (2002).
- Fishelovitch, D., Shaik, S., Wolfson, H. J. & Nussinov, R. Theoretical characterization of substrate access/exit channels in the human cytochrome P450 3A4 enzyme: Involvement of phenylalanine residues in the gating mechanism. *J. Phys. Chem. B* **113**, 13018–13025 (2009).
- Pollegioni, L., Wels, G., Pilone, M. S. & Ghisla, S. Kinetic mechanisms of cholesterol oxidase from *Streptomyces hygroscopicus* and *Brevibacterium sterolicum*. *Eur. J. Biochem.* **264**, 140–151 (2001).
- Piubelli, L. *et al.* On the oxygen reactivity of flavoprotein oxidases - An oxygen access tunnel and gate in *Brevibacterium sterolicum* cholesterol oxidase. *J. Biol. Chem.* **283**, 24738–24747 (2008).
- Setoyama, C. *et al.* Effects of hydrogen bonds in association with flavin and substrate in flavoenzyme D-amino acid oxidase. The catalytic and structural roles of Gly313 and Thr317. *J. Biochem.* **131**, 59–69 (2002).
- Sheng, D., Ballou, D. P. & Massey, V. Mechanistic studies of cyclohexanone monooxygenase: Chemical properties of intermediates involved in catalysis. *Biochemistry* **40**, 11156–11167 (2001).

33. Valley, M. P., Fenny, N. S., Ali, S. R. & Fitzpatrick, P. F. Characterization of active site residues of nitroalkane oxidase. *Bioorganic Chemistry* **38**, 115–119 (2010).
34. Wu, X., Palfey, B. A., Mossine, V. V. & Monnier, V. M. Kinetic studies, mechanism, and substrate specificity of amadoriase I from *Aspergillus* sp. *Biochemistry* **40**, 12886–12895 (2001).
35. Ruiz-Dueñas, F. J., Ferreira, P., Martínez, M. J. & Martínez, A. T. *In vitro* activation, purification, and characterization of *Escherichia coli* expressed aryl-alcohol oxidase, a unique H₂O₂-producing enzyme. *Protein Express. Purif.* **45**, 191–199 (2006).
36. Olsson, M. H. M., Sondergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
37. Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design* **27**, 221–234 (2013).
38. Gordon, J. C. *et al.* H⁺: a server for estimating pK_as and adding missing hydrogens to macromolecules. *Nucleic Acids Research* **33**, W368–W371 (2005).
39. Jacobson, M. P. *et al.* A hierarchical approach to all-atom protein loop prediction. *Proteins* **55**, 351–367 (2004).
40. Jacobson, M. P., Friesner, R. A., Xiang, Z. & Honig, B. On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* **320**, 597–608 (2002).
41. Bowers, K. J. *et al.* Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proc. ACM/IEEE Conference on Supercomputing (SC06)*, Tampa, 11–17 November 43 (2006).
42. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. 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 (2001).
43. Toukan, K. & Rahman, A. Molecular-dynamics study of atomic motions in water. *Phys. Rev. B* **31**, 2643–2648 (1985).
44. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
45. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant-Pressure Molecular-Dynamics Algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).
46. Tuckerman, M., Berne, B. J. & Martyna, G. J. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics* **97**, 1990–2001 (1992).
47. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577–8593 (1995).
48. Murphy, R. B., Philipp, D. M. & Friesner, R. A. A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *J. Comp. Chem.* **21**, 1442–1457 (2000).
49. Bochevarov, A. D. *et al.* Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *International Journal of Quantum Chemistry* **113**, 2110–2142 (2013).
50. Atilgan, A. R. *et al.* Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical Journal* **80**, 505–515 (2001).

Acknowledgements

This work was supported by the EnzOx2 project (H2020-BBI-PPP-2015-720297) of the European Joint Undertaking of Bio-based Industries (<http://bbi-europe.eu>), the INDOX project (KBBE-2013-7-613549) of the European Seventh Framework Programme, and the NOESIS (BIO2014-56388-R), vMutate (CTQ2016-79138-R) and FLADIMOTEC (BIO2016-75183-P) projects of the Spanish Ministry of Economy and Competitiveness. J.C. acknowledges a FPU fellowship (FPU2012-2041) from the Spanish Ministry of Education, Culture and Sports.

Author Contributions


J.C. performed the experiments and wrote the experimental sections of the paper. P.A.-R. and F.S. carried out the computational simulations and wrote the computational sections of the manuscript. A.T.M., P.F. and M.M. designed and supervised the experimental studies. V.G. designed and supervised the computational studies. A.T.M., P.F., V.G. and M.M. supervised the discussion of results and main conclusions. All authors revised and approved the submitted manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-26445-x>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

Multiple implications of an active site phenylalanine in the catalysis of aryl-alcohol oxidase

Juan Carro, Pep Amengual-Rigo, Ferran Sancho, Milagros Medina, Victor Guallar, Patricia Ferreira, Angel T. Martínez

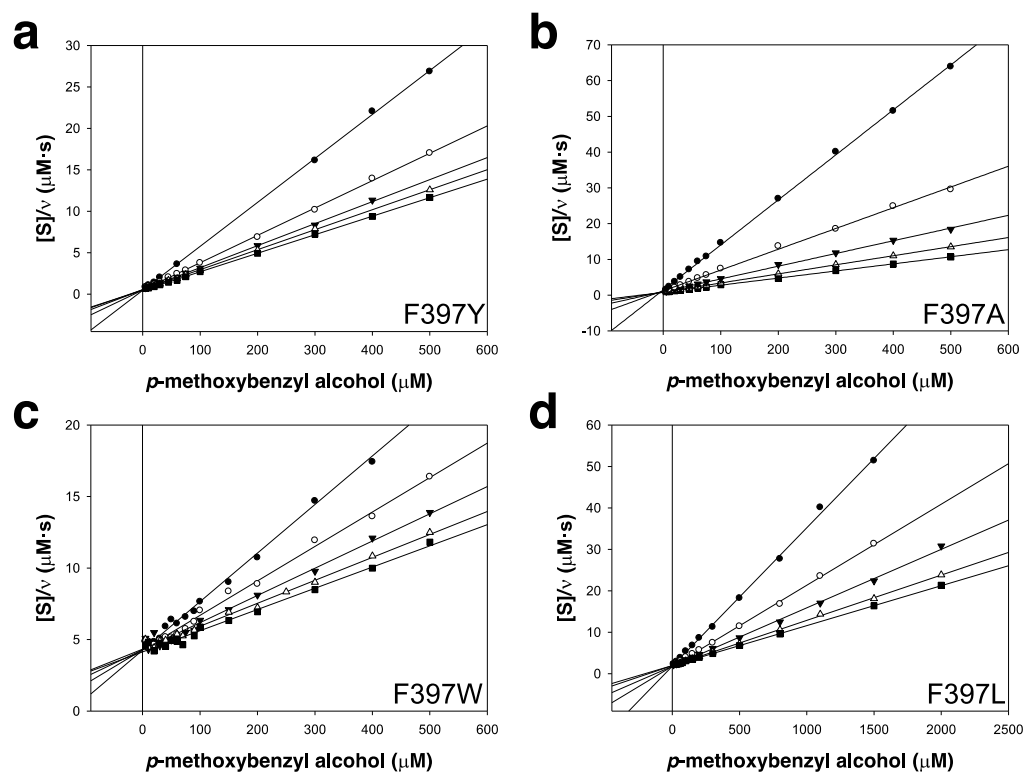
Supplementary Information

Supplementary information is composed of Supplementary Table S1 and Supplementary Figures S1, S2, S3, S4, S5 and S6.

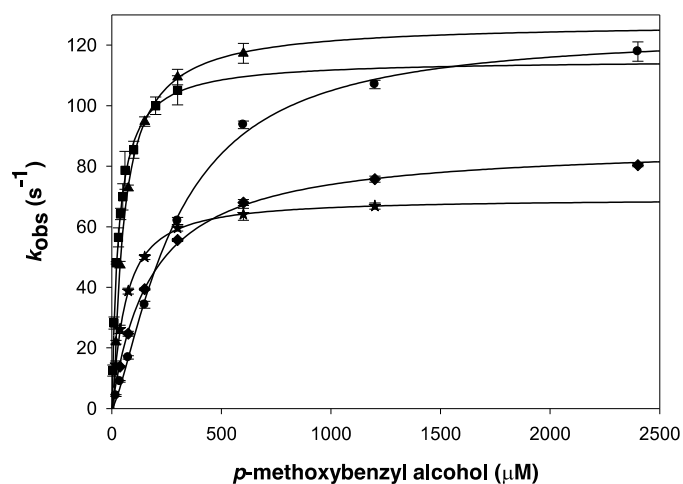
Supplementary table S1. Comparison of reaction constants measured as *p*-anisaldehyde or H₂O₂ production for Phe397 AAO variants at 25°C

	<i>p</i> -anisaldehyde			H ₂ O ₂		
	k_{cat} (s ⁻¹)	K_{m} (μM)	$k_{\text{cat}}/K_{\text{m}}$ (s ⁻¹ ·mM ⁻¹)	k_{cat} (s ⁻¹)	K_{m} (μM)	$k_{\text{cat}}/K_{\text{m}}$ (s ⁻¹ ·mM ⁻¹)
AAO ¹	113 ± 2	30 ± 2	3770 ± 260	106 ± 5	33 ± 6	3212 ± 603
F397Y	77 ± 2	23 ± 2	3400 ± 300	71 ± 2	23 ± 2	3060 ± 310
F397W	127 ± 6	123 ± 17	1030 ± 150	124 ± 4	125 ± 11	995 ± 94
F397A	54 ± 1	35 ± 3	1560 ± 140	48 ± 1	69 ± 4	698 ± 40
F397L	97 ± 6	155 ± 28	624 ± 120	92 ± 2	173 ± 17	530 ± 53

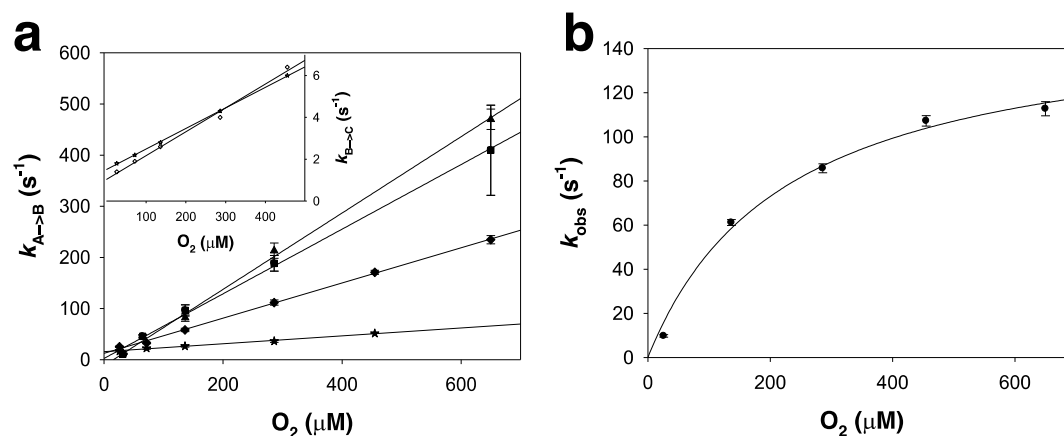
Reactions performed in 50 mM sodium phosphate pH 6.0 at 25°C and at a fixed O₂ concentration of 0.26 mM. *p*-Anisaldehyde release measured as increase of absorbance ($\Delta\epsilon_{285} = 16950 \text{ M}^{-1}\cdot\text{cm}^{-1}$). H₂O₂ release measured through a coupled enzymatic reaction with horseradish peroxidase (0.5 mM) and AmplexRed[®] (0.35 mM) measured as increase in absorbance ($\Delta\epsilon_{563} = 52000 \text{ M}^{-1}\cdot\text{cm}^{-1}$).¹From Ferreira et al.²². Means and standard deviations calculated from the fit to Michaelis-Menten equation. All kinetics were measured by triplicates.



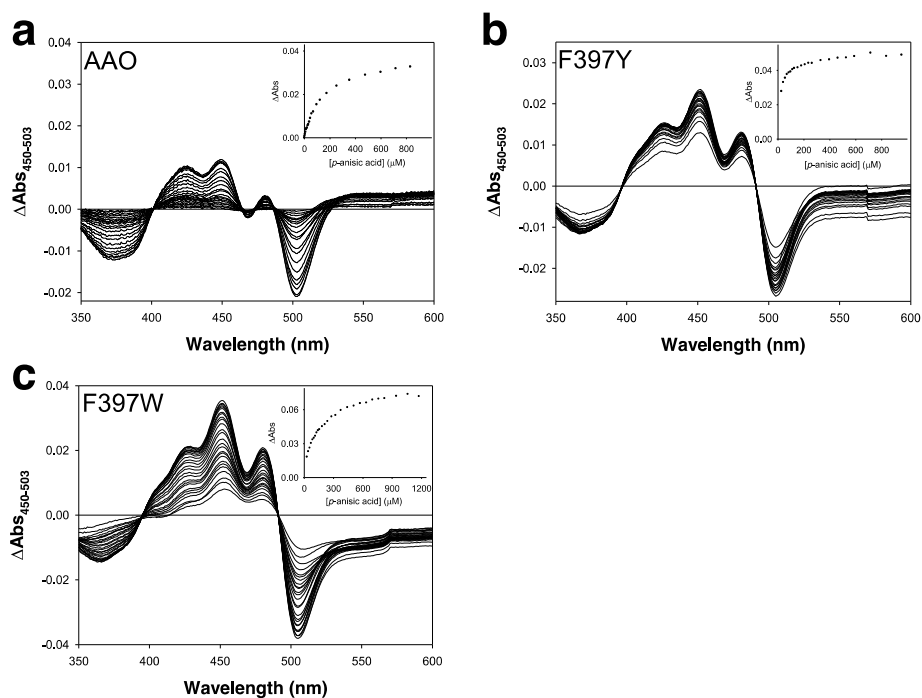
Supplementary figure S1. Linearized Hanes-Woolf plots of the bi-substrate kinetics of the Phe397 AAO variants. A. F397Y, B. F397A, C. F397W and D. F397L. Data were measured by varying the concentrations of *p*-methoxybenzyl alcohol and O₂ in 50 mM sodium phosphate pH 6.0 at 12°C. Vertical solid line represents $x = 0$. Filled circles, 0.06 mM O₂; open circles, 0.16 mM O₂; filled triangles, 0.34 mM O₂; open triangles, 0.71 mM O₂; and filled squares, 1.60 mM O₂.



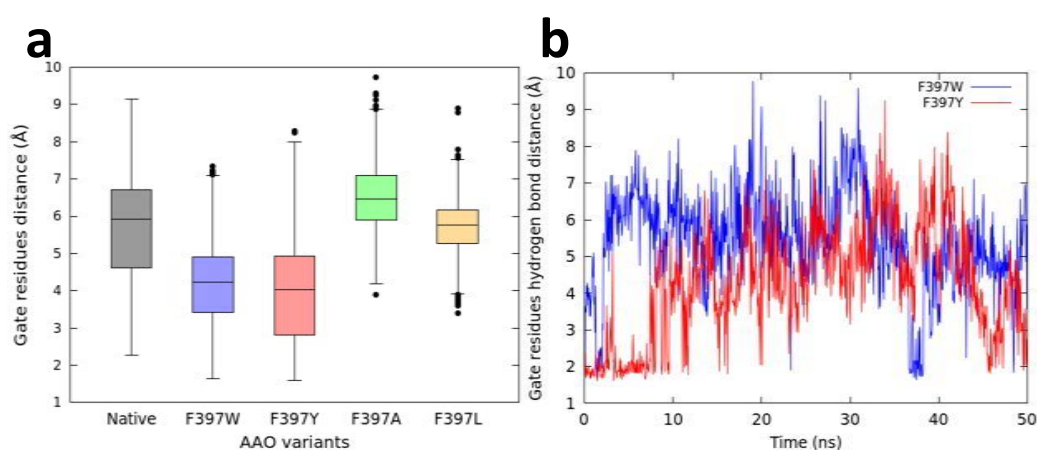
Supplementary figure S2. Plots of the observed rate constants (k_{obs}) of flavin reduction for native AAO and the four Phe397 variants. Native (squares), F397Y (triangles), F397W (circles), F397L (diamonds) and F397A (stars) AAO variants with varying concentrations of *p*-methoxybenzyl alcohol. Data were measured in an anaerobic stopped-flow spectrophotometer in 50 mM sodium phosphate pH 6.0 at 12°C. Lines represent the fits of experimental data to equation (3). Means and standard deviation calculated from triplicates.



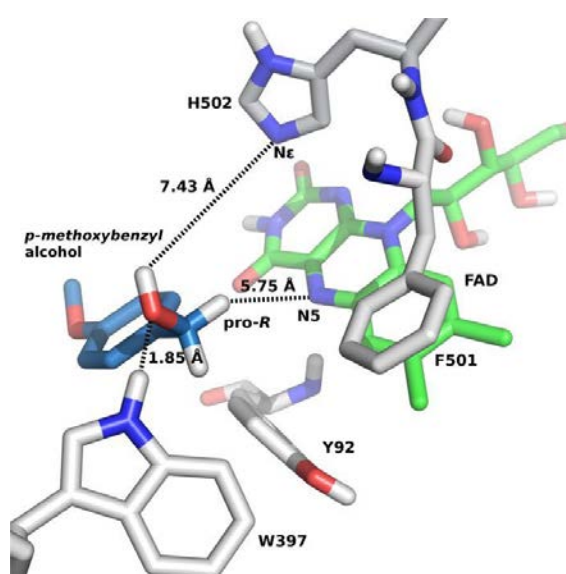
Supplementary figure S3. Plots of the observed rate constants (k_{obs}) of flavin oxidation as a function of O_2 concentration for native AAO and the four Phe397 variants. A. Native (squares), F397Y (triangles), F397A (stars) and F397L (diamonds). Inset shows the O_2 -dependent k_{obs2} for the second phase of the reoxidation of F397A (open stars) and F397L (open diamonds). Lines show fit to equation (4). **B.** F397W as a function of O_2 concentration. Line shows fit to equation (5). Data were measured in a stopped-flow spectrophotometer under anaerobic conditions at 12°C and pH 6.0. Means and standard deviation estimated from triplicates.



Supplementary figure S4. Spectral changes upon titration of native and Phe397 AAO variants with increasing concentrations of *p*-anisic acid. Enzyme (~10 μM , initial concentration) was titrated with *p*-anisic acid (0–1200 μM) at 25°C and pH 6.0. Spectra of the ligand-free enzyme and after each addition of ligand were recorded. All spectra were subtracted to that of the titrated enzymes to obtain the data represented. **A.** Native AAO, **B.** F397Y, and **C.** F397W variants, respectively. Insets show the differences in absorbance between maxima (450 nm) and minima (503 nm) as a function of the concentration of ligand.



Supplementary figure S5. Distance between gate-residues (residue at 397 position and Tyr92) computed by molecular dynamics of the Gln395–Thr406 and Ser89–Met95 loops. **A.** Boxplots of the distances. Boxes contain 50% of the results (two quartiles), horizontal lines inside them indicate the mean value; while the upper and lower whiskers (vertical lines) contain the remaining quartiles (50%). The overall height of boxes (whiskers included) is indicative of the spread of the results. Isolated points represent results significantly different from the rest of data values. Grey, native AAO; blue, F397W; red, F397Y; yellow, F397A; and light green, F397L variants. **B.** Representation of the evolution of the distances as a function of time (0–50 ns) for F397W and F397Y variants.



Supplementary figure S6. Representation of the energetically-favourable inefficient catalytic pose of *p*-methoxybenzyl alcohol in the F397W variant. Distances to His502, Trp397 and N5 of the FAD are indicated. Sticks in CPK, carbons in light blue.

1.2. Article II

Switching the substrate preference of fungal aryl-alcohol oxidase: towards stereoselective oxidation of secondary benzyl alcohols

Ana Serrano,^{‡a} Ferran Sancho,^{‡b} Javier Viña-González,^c Juan Carro,^a Miguel Alcalde,^c 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 Barcelona Supercomputing Center, Jordi Girona 31, E-08034 Barcelona, Spain

^c Department of Biocatalysis, Institute of Catalysis, CSIC, Cantoblanco, E-28049, Madrid, Spain

^d Icrea, Passeig Lluís Companys 23, E-08010, Barcelona, Spain

[‡] These two authors equally contributed to the work.



Cite this: *Catal. Sci. Technol.*, 2019, 9, 833

Switching the substrate preference of fungal aryl-alcohol oxidase: towards stereoselective oxidation of secondary benzyl alcohols†

Ana Serrano,^{‡a} Ferran Sancho,^{‡b} Javier Viña-González,^c Juan Carro,^a Miguel Alcalde,^c Victor Guallar^{bd} and Angel T. Martínez^{†*a}

Oxidation of primary alcohols by aryl-alcohol oxidase (AAO), a flavoenzyme that provides H₂O₂ to fungal peroxidases for lignin degradation in nature, is achieved by concerted hydroxyl proton transfer and stereoselective hydride abstraction from the pro-*R* benzylic position. In racemic secondary alcohols, the *R*-hydrogen abstraction would result in the selective oxidation of the *S*-enantiomer to the corresponding ketone. This stereoselectivity of AAO may be exploited for enzymatic deracemization of chiral mixtures and isolation of *R*-enantiomers of industrial interest by switching the enzyme activity from primary to secondary alcohols. A combination of computational simulations and mutagenesis has been used to produce AAO variants with increased activity on secondary alcohols, using the already available F501A variant of *Pleurotus eryngii* AAO as a starting point. Adaptive-PELE simulations for the diffusion of (*S*)-1-(*p*-methoxyphenyl)-ethanol in this variant allowed Ile500 to be identified as one of the key residues with a higher number of contacts with the substrate during its transition from the solvent to the active site. Substitution of Ile500 produced more efficient variants for the oxidation of several secondary alcohols, and the I500M/F501W double variant was able to fully oxidize (after 75 min) with high selectivity (ee >99%) the *S*-enantiomer of the model secondary aryl-alcohol (±)-1-(*p*-methoxyphenyl)-ethanol, while the *R*-enantiomer remained unreacted.

Received 3rd December 2018,
Accepted 15th January 2019

DOI: 10.1039/c8cy02447b

rsc.li/catalysis

Introduction

There is an increasing interest in the production of pure enantiomers of a variety of chemical compounds for preparation of drugs and fine chemicals.¹ The majority of chiral molecules of industrial interest are obtained by kinetic resolution from racemic mixtures in which two enantiomers react at different rates with a chiral catalyst resulting in a sample enantio-enriched with the less reactive isomer.² For this purpose, the use of biological systems (microorganisms and enzymes) provides an alternative to chemical reagents due to the regio- and enantio-selectivity of many biocatalysts and their mild reaction conditions.^{3–6}

For oxidation of secondary alcohols in an enantioselective manner, both dehydrogenases and oxidases have been

used.^{3,4,7} Due to the intrinsic asymmetry of these enzymes, their action on secondary alcohols often results in kinetic resolution with selectivity and enantiomeric excesses (ee) depending on the characteristics of their active sites. Several NAD(P)H-dependent alcohol dehydrogenases have been described for deracemization of secondary alcohols.^{8,9} However their use implies the need for auxiliary enzymes and stoichiometric amounts of reductants to recycle the NAD(P)H co-substrate. Thus, oxidases, which only need molecular oxygen as an oxidizing agent, are an interesting option for the stereoselective oxidation of secondary alcohols.⁶ Polyvinyl-alcohol oxidase is described as a secondary alcohol oxidase, although no information on its eventual selectivity is available.¹⁰ Other oxidases have been reported to oxidize secondary alcohols,^{4,6} including cholesterol oxidase,¹¹ glycolate oxidase¹² and alditol oxidase.¹³ Moreover, due to the importance of stereoselective oxidation of these substrates, several studies that widen the specificity of other oxidases on secondary alcohols have been reported in the last few years.^{14,15}

One potential candidate for oxidation of secondary alcohols is fungal aryl-alcohol oxidase from *Pleurotus eryngii* (AAO, E.C 1.1.3.7), whose biotechnological potential has been demonstrated.^{16,17} This flavooxidase catalyzes the oxidation of a range of primary alcohols conjugated to an aromatic

^a Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain. E-mail: atmartinez@cib.csic.es

^b Barcelona Supercomputing Center, Jordi Girona 31, E-08034 Barcelona, Spain

^c Department of Biocatalysis, Institute of Catalysis, CSIC, Cantoblanco, E-28049, Madrid, Spain

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

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8cy02447b

‡ These two authors equally contributed to the work.



group (mainly phenyl but also naphthalenyl and furanyl) or even to an aliphatic-polyunsaturated system.¹⁸ The AAO catalytic mechanism consists of proton transfer from the hydroxyl group to a catalytic base, His502, taking place in a concerted (but asynchronous) way with hydride abstraction from the benzylic position by the oxidized flavin.^{19,20} Due to its active-site architecture and the concerted nature of the hydride and proton transfers, hydride abstraction in AAO is stereoselective, taking place only from the pro-*R* position.²¹ This stereoselectivity could be exploited for deracemization since it is maintained when secondary benzyl alcohols (with chiral centers) are assayed as substrates, although the AAO activity on these compounds is almost residual due to its narrow active site. In fact, it has been reported that the widening of the active site by substitution of the bulky Phe501 in variant F501A increases the enzyme activity on secondary alcohols, and improves its stereoselectivity.²¹

In this work, we took advantage of computational simulations to design AAO variants with increased activity on secondary alcohols. These variants were kinetically characterized to evaluate their oxidation efficiency (on primary and secondary benzyl alcohols) and their selectivity for deracemization reactions (Scheme 1) was assessed. A switch in the substrate pattern of the variants was observed and rationalized at the atomic level.

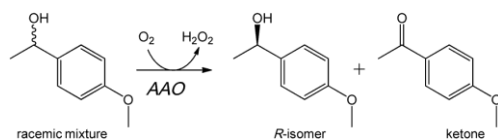
Material and methods

Chemicals

p-Methoxybenzyl alcohol, (\pm)-1-(*p*-methoxyphenyl)-ethanol (racemic mixture), (*R*)-1-(*p*-fluorophenyl)-ethanol, (*S*)-1-(*p*-fluorophenyl)-ethanol, (\pm)-1-phenylpropanol (racemic mixture), 2-phenyl-2-propanol and horseradish peroxidase (HRP) were purchased from Sigma-Aldrich. (\pm)-2-Methyl-1-phenyl-1-propanol (racemic mixture) was purchased from Alfa Aesar. AmplexRed® was obtained from Invitrogen.

Directed mutagenesis, enzyme production and purification

Wild-type recombinant (hereinafter native) AAO from *P. eryngii* was obtained by expressing the mature AAO cDNA (GenBank AF064069) in *Escherichia coli* followed by *in vitro* activation and purification, as previously described.²² Variants were produced by site-directed mutagenesis using the following synthetic primers (bold substituted nucleotides in underlined mutated triplets are indicated below): 5'-GGG TCT AGC TCT GTT CAC TTC ATG GTC ATG ATG CG-3' for



Scheme 1 Stereoselective oxidation of a racemic mixture of model secondary benzyl alcohol (1-[*p*-methoxyphenyl]-ethanol) by AAO leading to kinetic resolution.

Y92F, 5'-GAC AAC GCC AAC ACG GCT TTC CAC CCA GTT GG-3' for I500A, 5'-GAC AAC GCC AAC ACG ATG TTC CAC CCA GTT GG-3' for I500M, 5'-GAC AAC GCC AAC ACG TGG TTC CAC CCA GTT GG-3' for I500W, 5'-C AAC GCC AAC ACG ATT GCC CAC CCA GTT GGA ACG-3' for F501A and 5'-GAC AAC GCC AAC ACG ATT TGG CAC CCA GTT GG-3' for F501W, using the plasmid containing the native AAO sequence as template. The double mutations were obtained using the I500A plasmid as a template and the primers 5'-C CTA TCC GAC CAT TTG GCC CTT CCT GCT GCC TTC TTC G-3' for L315A/I500A and 5'-GCT CAT TGG GAG ACC GCC TTT TCC AAC CAA TGG-3' for I391A/I500A and the F501W plasmid as a template and the primer 5'-GAC AAC GCC AAC ACG ATG TGG CAC CCA GTT GG-3' for I500M/F501W including both mutations. Mutations were confirmed by gene sequencing, the variants were purified to electrophoretic homogeneity following the same protocol as for the native protein,²² and their electronic absorption spectra were recorded. Enzymes were quantified with a Cary-4000 spectrophotometer using extinction coefficients (Table S1†) taken from the literature or calculated in the present work by heat denaturation,²³ using $\epsilon_{450} = 11\,300\text{ M}^{-1}\text{ cm}^{-1}$ for the free FAD.²⁴

Steady-state kinetics for alcohol oxidation

The kinetic parameters for oxidation of primary *p*-methoxybenzyl alcohol were calculated by following spectrophotometrically the oxidation initial rate (1 min reaction) of the alcohol to the corresponding aldehyde using the difference molar absorptivity ($\Delta\epsilon_{285} = 16\,950\text{ M}^{-1}\text{ cm}^{-1}$).²⁵

The kinetic parameters for oxidation of the secondary alcohol (\pm)-1-(*p*-methoxyphenyl)-ethanol were calculated by monitoring the production of H₂O₂ in the HRP-coupled assay with AmplexRed ($\Delta\epsilon_{563} = 52\,000\text{ M}^{-1}\text{ cm}^{-1}$) as an alternative to the spectrophotometric estimation of the ketone product.

Measurements were performed at 25 °C in air-saturated (0.256 mM O₂ concentration)²⁶ 50 mM phosphate, pH 6.0. Kinetic parameters were determined by fitting the initial reaction rates at different alcohol concentrations to the Michaelis-Menten equation (eqn (1)):

$$\frac{v}{e} = \frac{k_{\text{cat}}[S]}{K_m + [S]} \quad (1)$$

In the case of (\pm)-1-(*p*-methoxyphenyl)-ethanol, the enantiomer concentration was considered to be 50% of the racemic mixture concentration, and an apparent turnover number ($^{\text{app}}k_{\text{cat}}$) was estimated in the presence of both enantiomers.

Analysis of stereoselective reactions

The time course of mid-term (up to 75 min) and long-term (up to 24 or 72 h) reactions of native AAO and variants with (\pm)-1-(*p*-methoxyphenyl)-ethanol, (\pm)-1-phenyl-1-propanol and (\pm)-2-methyl-1-phenyl propanol was followed at 25 °C in air saturated 50 mM phosphate, pH 6.0, under continuous shaking. The samples were taken at different times and, after



addition of an internal standard (Table S2†), were liquid–liquid extracted with hexane, and analyzed by chiral HPLC.

The enantiomers were separated in a Chiralcel IB column (4.6 × 250 mm, 5 μm; Daicel Chemical Industries, Ltd.) using a pre-column of the same material, and 98 : 2 (v/v) *n*-hexane : isopropanol (98 : 1 for 2-methyl-1-phenyl propanol) as a mobile phase (at 1 mL min⁻¹ and 25 °C). Detection was performed at 206 nm (see spectra of the different compounds in Fig. S1†). The retention times for the corresponding *R* and *S* enantiomers (Table S2†) were obtained from the racemic standards assuming the published elution order.²¹ Quantification was performed using internal standards and calibration curves (Fig. S2†).

The ee was calculated using eqn (2):

$$ee = 100 \times \left(\frac{R - S}{R + S} \right) \quad (2)$$

where *R* and *S* are the amounts of each enantiomer calculated from the calibration curves.

Enantiomeric ratios (*E*-values)²⁷ were calculated from eqn (3):

$$E = \frac{\ln \left[\frac{(1-C)(1-ee)}{(1-C)(1-ee)} \right]}{\ln \left[\frac{(1-C)(1-ee)}{(1-C)(1-ee)} \right]} \quad (3)$$

where *C* is the conversion rate.

The oxidation of (*R*)- and (*S*)-1-(*p*-fluorophenyl)-ethanol to *p*-fluoroacetophenone by native AAO and variants was evaluated in air-saturated 50 mM phosphate, pH 6.0, at 25 °C, after long-term incubations (up to 48 h). The amount of *p*-fluoroacetophenone formed was calculated using the difference in the molar absorbance coefficients of the alcohol and the ketone at 248 nm ($\epsilon_{\text{ketone}} - \epsilon_{\text{alcohol}}$, $\Delta\epsilon_{248} = 12\,606 \text{ M}^{-1} \text{ cm}^{-1}$) (Fig. S3†).

Computational analysis

The new adaptive-PELE (protein energy landscape exploration) software²⁸ was used to study (*R*)- and (*S*)-1-(*p*-methoxyphenyl)-ethanol diffusion and binding on native AAO and four variants. The adaptive protocol improves sampling in PELE by running multiple short simulations (epochs) where the initial conditions in each of them are selected through a reward function aiming at sampling non-visited areas. Briefly, PELE uses a Monte Carlo (MC) procedure including protein structure prediction algorithms and a ligand rotamer library for sampling enhancement. Each MC iteration includes three main steps: 1) ligand and protein (backbone) perturbation; 2) side chain sampling; 3) overall minimization. For this study, ligand perturbation involved the [0.5–1.5] translation and [0.05–0.1] rotation ranges, in Å and rad, respectively. Backbone flexibility was allowed using the lowest 6 modes in an anisotropic network model,²⁹ while all side chains within 6 Å of the ligand were predicted on each step. The ligand was allowed to move enforcing (its center of mass) a 5 Å radius sphere center on the FAD N5 coordinates. Each

simulation involved 192 trajectories with 40 epochs and 20 MC PELE iterations per epoch. To improve the sampling towards the FAD cofactor, we used an epsilon value of 0.1, meaning that 10% of the processors started each epoch from the best ligand-FAD distance previously sampled. More details and examples on running PELE can be found elsewhere.³⁰

PELE uses an all atom OPLS2005 force field,³¹ with an implicit generalized born model. All ligands and FAD charges, however, were extracted from quantum mechanical (QM) calculations. The FAD cofactor (in its quinone state) was optimized with mixed quantum mechanics/molecular mechanics (QM/MM) calculations at the M06(6-31G*)/OPLS2005 level of theory using Qsite.³² The initial model was derived from the 3FIM crystal structure, solvated with an 8 Å layer of water molecules and prepared with Maestro's protein wizard.³³ Only the FAD, which does not present any covalent interaction with the rest of the protein, was included in the QM region. Ligands were optimized at the same QM level of theory with an implicit PBF solvent using Jaguar.³⁴ Ligands were then parameterized in accordance to OPLS2005, keeping the electrostatic QM charges and a rotamer library was built with MacroModel.³³

Results and discussion

Although the activity of AAO on secondary alcohols is nearly residual, a variant with increased activity was obtained by substituting the bulky Phe501 by an alanine.²¹ Taking this variant as a starting point, we combined computational simulations and site-directed mutagenesis to obtain new variants with higher activity on secondary alcohols of interest in deracemization reactions.

Computational approach for AAO engineering

To identify mutations along the access path to the AAO active site (Fig. 1A) that could improve the secondary alcohol access and binding, PELE simulations of (*S*)-1-(*p*-methoxyphenyl)-ethanol diffusion in the F501A variant (molecular structure from *in silico* mutation of PDB 3FIM) were carried out. The goal was to identify residues at the active-site access channel potentially-involved in the substrate diffusion and oxidation. The strategy included exploring the energy profiles for the entrance of this secondary alcohol and determining which amino acids can be limiting the ligand's access or constraining it in an incorrect orientation for oxidation at the active site.

During the ligand transition from the solvent to the active site, several residues showed significantly higher number of interaction contacts with the alcohol (Fig. 2) being the main direct obstacles that the substrate has to bypass. Among them, Tyr92 and Phe397 together with Phe501 in native AAO constitute a hydrophobic bottleneck (Fig. 1A) for the access of substrates, being also involved in their stabilization at the active site.^{35–38} According to these computational data, residues with contact numbers higher than 35 000—such as Pro79, Tyr92, Leu315, Ile391, Phe397, Pro399 and Ile500, together with



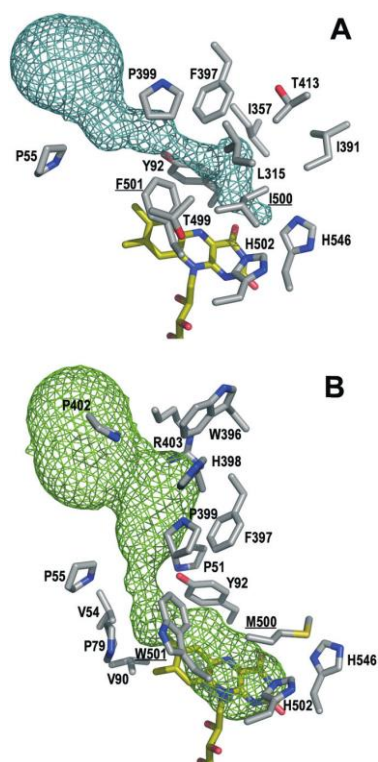


Fig. 1 Detail of the access channel and active site cavity (around flavin N5) in native AAO (A) and its regioselective I500M/F501W variant (B), including FAD (yellow-C CPK-colored sticks) and residues at less than 4 Å from the channel (side chains as gray-C CPK-colored sticks; mutated residues underlined).

previously mutated Ala501, all of them located at less than 4 Å of the active-site channel—were selected for site-directed mutagenesis. *In vitro* folding of the mutated proteins was performed after replacing the above residues with alanines, and alternative mutations were introduced in several cases.

Proper incorporation of the FAD cofactor was shown by the presence of typical bands I and II in their electronic absorption spectra (with only slight displacements due to mutations in the flavin environment; Table S1†). The ability of these variants to oxidize secondary alcohols was tested by incubating them with (\pm)-1-(*p*-methoxyphenyl)-ethanol, as a model chiral substrate, and analyzing the resulting ketone and the remaining *R* and/or *S* substrate enantiomers (Fig. S2†). Quantification of both isomers by chiral HPLC after 24 h of reaction indicated that, with the exception of I500W, the variants oxidized (*S*)-1-(*p*-methoxyphenyl)-ethanol to the corresponding ketone with different rates, without any activity on (*R*)-1-(*p*-methoxyphenyl)-ethanol. I500A showed 15-times higher activity than the native enzyme, with 50% conversion of the racemic substrate in the first 4 h of reaction (Fig. 3 and Table S3†) due to the almost total oxidation of the *S*-enantiomer.

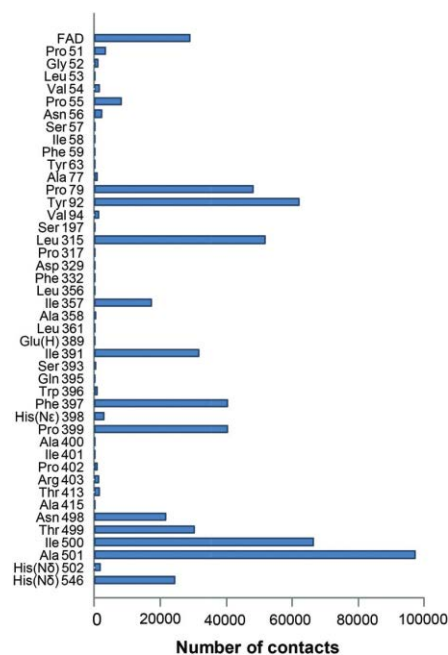


Fig. 2 Residues (up to a total of 43 plus the FAD cofactor) interacting with the model secondary alcohol (*S*)-1-(*p*-methoxyphenyl)-ethanol during PELE simulations (on the F501A variant). Interactions were selected considering contact distances below 2.7 Å between the ligand and enzyme atoms. See Fig. 1 for the location of the main residues. Glu(H) indicates a protonated glutamate, and His(N ϵ) and His(N δ) indicate histidines protonated in N ϵ and N δ , respectively.

Rationalizing the effect of the I500A mutation

PELE simulations for (*S*)-1-(*p*-methoxyphenyl)-ethanol diffusion and positioning at the active site of the I500A variant, compared to native AAO, contributed to explaining its higher activity on secondary alcohols mentioned above.

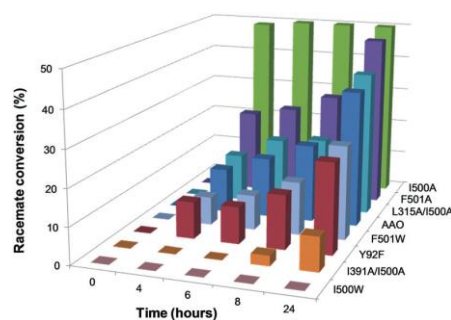


Fig. 3 Conversion of (\pm)-1-(*p*-methoxyphenyl)-ethanol during 24 h reaction with native AAO and seven variants. The reactions between the alcohol (2.5 mM racemic mixture) and the enzyme (5 μ M) were performed in 50 mM phosphate, pH 6.0, at 25 °C, and analyzed by HPLC.



In the case of native AAO, PELE diffusion leads to two main locations of the alcohol in the active site, in which its hydroxyl group is located near the catalytic His502 at an adequate distance (~ 2.5 Å) for the proton abstraction to be produced (Fig. 4A). These two structures differ in the position of the benzylic hydrogen (in the *R* position) with regard to the flavin (hydrogen–FAD distances of 2.5 Å and 5.3 Å) only the first one being compatible with hydride transfer to flavin N5 and ketone formation.

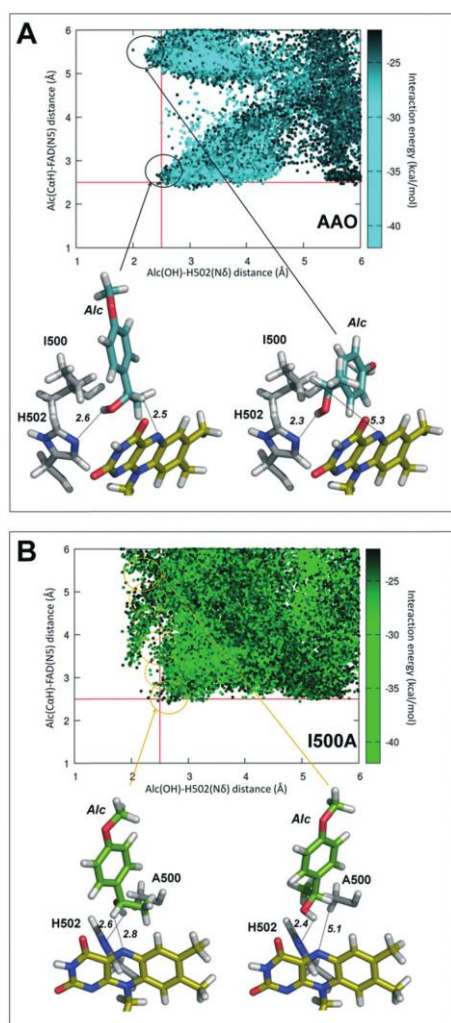


Fig. 4 PELE diffusion of (*S*)-1-(*p*-methoxyphenyl)-ethanol (A1c) on native AAO (A) and its I500A variant (B) including: i) plot of the PELE results (top) with each position represented as a function of the distances between the hydroxyl hydrogen and the N δ of His502 (axis-*x*) and the benzylic hydrogen (C α) and the flavin N5 (axis-*y*); and ii) two selected substrate poses (bottom) for each simulation, including the substrate, flavin ring, catalytic His502 and the residue at the mutated position. Each point is colored according to the interaction energy scale.

Table 1 Kinetic constants of native AAO and four selected variants on model primary (top) and secondary (bottom) benzyl alcohols^a

<i>p</i> -Methoxybenzyl alcohol			
	k_{cat} (s ⁻¹)	K_{m} (μ M)	$k_{\text{cat}}/K_{\text{m}}$ (mM ⁻¹ s ⁻¹)
AAO	142 \pm 5	27 \pm 4	5230 \pm 620
F501A	3.1 \pm 0.03	12.3 \pm 0.6	251 \pm 12
I500A	1.5 \pm 0.1	1.3 \pm 0.1	1140 \pm 45
I500M	5.3 \pm 0.3	0.4 \pm 0.1	13 100 \pm 200
I500M/F501W	3.3 \pm 0.1	3.1 \pm 0.03	7930 \pm 640
(\pm) 1-(<i>p</i> -Methoxyphenyl)-ethanol			
	^{app} k_{cat} (s ⁻¹) ^b	K_{m} (mM) ^c	^{app} $k_{\text{cat}}/K_{\text{m}}$ (mM ⁻¹ s ⁻¹)
AAO	0.18 \pm 0.002	25.0 \pm 0.6	0.0072 \pm 0.0002
F501A	0.05 \pm 0.003	10.0 \pm 1.5	0.0051 \pm 0.0008
I500A	0.22 \pm 0.01	2.9 \pm 0.3	0.079 \pm 0.009
I500M	0.42 \pm 0.01	1.4 \pm 0.1	0.30 \pm 0.02
I500M/F501W	2.2 \pm 0.04	3.1 \pm 0.2	0.71 \pm 0.05

^a Determined at 25 °C in air-saturated (0.256 mM O₂) 50 mM phosphate, pH 6.0. ^b Apparent k_{cat} estimated in the racemic mixture. ^c Referring to the *S*-enantiomer, representing 50% of the racemic mixture.

By contrast, alcohol diffusion in I500A showed that in this variant the ligand is able to move more freely in the catalytic site (due to its increased size) with a range of positions (circles in Fig. 4B, top) at 2.0–2.5 Å from His502 and progressively shorter distances of the FAD, to finally attain a position compatible with catalysis (Fig. 4B, bottom, left).

Combinatorial saturation mutagenesis of AAO expressed in *Saccharomyces cerevisiae*,³⁹ taking the above results into account, yielded the I500M/F501W variant with a noticeable increase in the oxidation of 1-(*p*-methoxyphenyl)-ethanol.⁴⁰

Alcohol oxidation by I500 and Phe501 variants

Therefore, to get insights into the AAO stereoselective oxidation of secondary benzyl alcohols, two additional variants (I500M and I500M/F501W) were expressed in *E. coli*, *in vitro* activated and purified to homogeneity, showing correct folding and FAD incorporation (with only slight displacements in bands I and II due to mutations in the flavin environment; Table S1†).

We first evaluated the effect of four selected single and double substitutions of Ile500 and Phe501 on the AAO activity towards its preferred substrate, *p*-methoxybenzyl alcohol,

Table 2 Changes in the relative secondary (1-(*p*-methoxyphenyl)-ethanol) to primary (*p*-methoxybenzyl alcohol) turnover (s⁻¹) ratio of four selected variants referring to the native AAO (from Table 1 data)

	Secondary (sec; s ⁻¹)	Primary (pri; s ⁻¹)	Relative sec/pri ratio
AAO	0.18	142.0	1
F501A	0.05	3.1	13
I500A	0.22	1.5	116
I500M	0.42	5.3	63
I500M/F501W	2.2	3.3	526



by measuring the release of anisaldehyde (Table 1, top). Aldehyde estimation has been confirmed to provide similar kinetic constants to those obtained when the equimolecular

H_2O_2 release is followed.³⁶ The variants showed lower turnover numbers (k_{cat}) but increased affinity, as shown by lower K_{m} values, with regard to the native enzyme. Thus, the I500M/F501W and I500M mutations resulted in 1.5- and 2.5-fold more efficient oxidation of the primary alcohol, respectively.

Then, kinetic measurements of (\pm)-1-(*p*-methoxyphenyl)-ethanol oxidation were followed by H_2O_2 release (Table 1, bottom). Increased affinity for the secondary alcohol was reflected in the lower K_{m} values that the variants showed. As a consequence, I500A, I500M and I500M/F501W present 11-, 42- and 97-fold higher catalytic efficiencies ($k_{\text{cat}}/K_{\text{m}}$) than native AAO, respectively, in agreement with the results from 24 h experiments (Table S3†). Comparison of the action of the variants on secondary and primary alcohols was based on their turnover (k_{cat}) values, since in deracemization reactions in an industrial context substrates are used under saturated conditions and the rates are therefore independent of the K_{m} value.⁴¹ Interestingly, the substitution of Ile500 produced an increase of the activity on the model secondary alcohol at the expense of reducing its activity on the primary *p*-methoxybenzyl alcohol, as shown by kinetic comparison (up to 500-fold higher secondary/primary ratio for I500M/F501W) (Table 2). Therefore, the more active a variant is on secondary alcohols, the less active it is on primary alcohols, revealing an interesting switch in the substrate preference of AAO. The higher activity of the double variant agrees with a broader active site (Fig. 1B) enabling

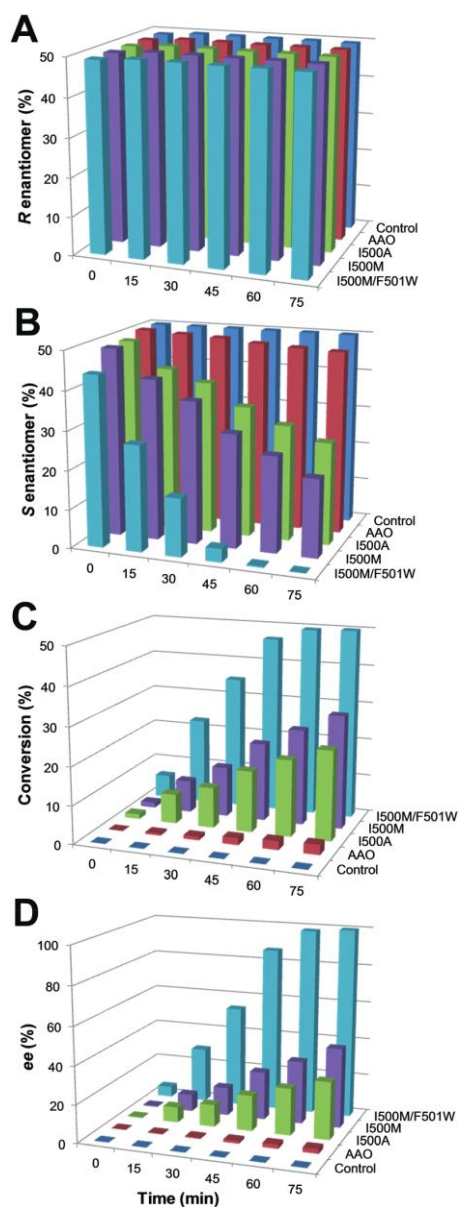


Fig. 5 (\pm)-1-(*p*-Methoxyphenyl)-ethanol oxidation to the corresponding ketone for 75 min by three selected AAO variants compared with the native enzyme (and control without enzyme): A) remaining *R*-enantiomer; B) remaining *S*-enantiomer; C) conversion yield (racemic mixture); and D) *R*-isomer ee. Reactions between alcohol (2.5 mM racemic mixture) and enzyme (2.5 μM) were performed in 50 mM phosphate, pH 6.0, at 25 $^{\circ}\text{C}$, and the remaining substrates and product were analyzed by chiral HPLC.

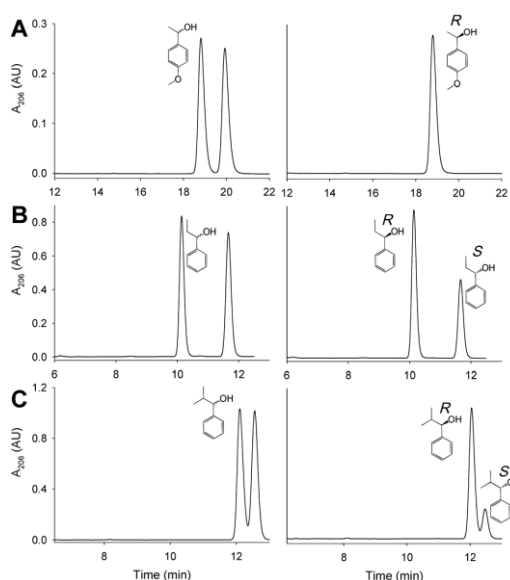


Fig. 6 Kinetic resolution of (\pm)-1-(*p*-methoxyphenyl)-ethanol (A), (\pm)-1-phenyl-1-propanol (B) and (\pm)-1-phenyl-2-methyl-1-propanol (C) by the I500M/F501W variant: chromatograms before (left) and after incubation with the enzyme (for 1 h in A and 72 h in B and C) resulting in *R*-enriched samples (right).



Table 3 Conversion yield and selectivity (ee and *E*-value of the *R*-enantiomer) in mid-term reactions of racemic (\pm)-1-(*p*-methoxyphenyl)-ethanol with native AAO and four selected variants^a

	Conversion ^b (%)	ee (%)	<i>E</i> -Value
AAO	3	2	— ^c
F501A	4	5	—
I500A	24	30	204
I500M	30	42	352
I500M/F501W	50	100	89 000

^a Mid-term (75 min) reaction of 2.5 mM alcohol and 2.5 μ M enzyme in air-saturated 50 mM phosphate, pH 6.0, at 25 °C. ^b Referring to the racemic mixture. ^c —, too low *E*-values (<15) for practical purposes.

entering and adequate positioning of the bulkier secondary alcohol.

Stereoselective oxidation of (\pm)-1-(*p*-methoxyphenyl)-ethanol

The oxidation of the model secondary aryl alcohol by native AAO and its variants was monitored in time-course reactions using chiral HPLC. Reactions were performed at 25 °C under continuous shaking and aliquots were taken every 15 min for 2 h (Fig. S4†). As shown in Fig. 5, only 3% conversion (with 2% ee) was found after 2 h incubation with native AAO, and the results were only slightly improved with the F501A variant.

However, other variants were considerably more active on the secondary alcohol, with up to 18- and 66-fold higher conversion rates in the cases of I500M and I500M/F501W, respectively (Fig. 5B and C). The substitution of Ile500 by either methionine or alanine considerably increases both conversion and ee with regard to native AAO. After 75 min, the conversion reached 24% and 30% for I500A and I500M (with ee of 30% and 42%), respectively. These values increased up to 50% conversion of the total racemic mixture (with ee 100%) for I500M/F501W, indicating a high enantioselectivity of this double variant (Fig. 5C and D, 6A and Table 3).

Reaction with other secondary alcohols

To extend the deracemization potential of AAO, revealed by the (\pm)-1-(*p*-methoxyphenyl)-ethanol reactions, to other secondary benzyl alcohols, the I500A, I500M and I500M/F501W vari-

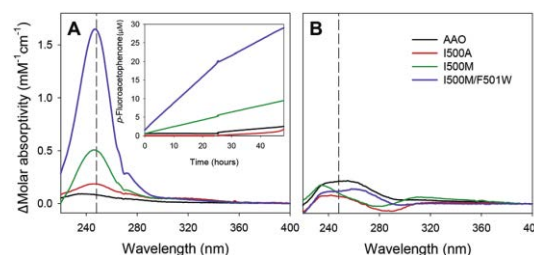


Fig. 7 Molar absorptivity difference spectra (final – initial) after 48 h of reaction of (*S*)-1-(*p*-fluorophenyl)-ethanol (A) and (*R*)-1-(*p*-fluorophenyl)-ethanol (B) with native AAO and three variants. The inset shows the *p*-fluoroacetophenone production. Reactions between alcohol (100 μ M) and enzyme (~0.2 μ M) were performed in 50 mM phosphate, pH 6.0, at 25 °C.

ants (10 μ M) were incubated with (\pm)-1-phenyl-1-propanol and (\pm)-1-phenyl-2-methylpropanol (2.5 mM) up to 72 h.

Chiral HPLC analyses indicated that I500M and I500M/F501W were able to oxidize the two alcohols to their corresponding ketones (Fig. 6B and C) although with low conversion yields (4% and 22% of the above racemic compounds, respectively, by I500M; and 13% and 31%, respectively, by I500M/F501W) leading to an *R*-enantiomer ee up to 62% in the case of I500M/F501W (Table 4). These results suggest that the active site in these two variants has been sufficiently enlarged (as illustrated in Fig. 1B for I500M/F501W) to accommodate not only the methyl group of the model secondary alcohol but also larger groups, such as the ethyl and isopropyl groups of the two other secondary alcohols assayed.

The enantioselectivity of the Ile500 variants was confirmed by incubating them with pure *R*- and *S*-enantiomers of 1-(*p*-fluorophenyl)-ethanol. After 48 h of reaction, the formation of *p*-fluoroacetophenone (up to 30%) was shown by difference spectra when (*S*)-1-(*p*-fluorophenyl)-ethanol was treated with the single and double Ile500 variants (Fig. 7A). By contrast, hardly any reaction was observed for (*R*)-1-(*p*-fluorophenyl)-ethanol (Fig. 7B) confirming the enantioselectivity shown using racemic mixtures.

Stereoselectivity explained by *R* and *S* simulations

The AAO selectivity for the *S*-enantiomers, during secondary alcohol oxidation, is consistent with the catalytic mechanism

Table 4 Conversion yield and selectivity (*R*-enantiomer ee) in long-term reaction of two additional secondary alcohols with native AAO and three selected variants^a

	1-Phenyl-1-propanol		1-Phenyl-2-methylpropanol	
	Conversion ^b (%)	ee (%)	Conversion ^b (%)	ee (%)
AAO	1	2	0	0
I500A	1	2	0	0
I500M	4	9	22	39
I500M/F501 W	13	26	31	62

^a Long-term (72 h) reaction of 2.5 mM alcohol and 10 μ M enzyme in air-saturated 50 mM phosphate, pH 6.0, at 25 °C. ^b Referring to the racemic mixture.



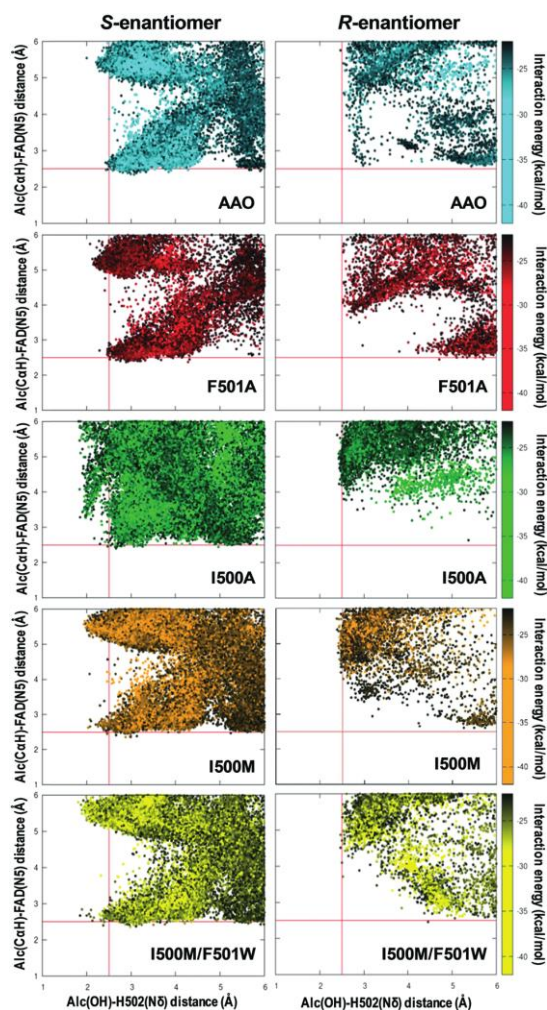


Fig. 8 PELE diffusion of (*S*)-1-(*p*-methoxyphenyl)-ethanol (left) and (*R*)-1-(*p*-methoxyphenyl)-ethanol (right) on AAO (blue) and its F501A (red), I500A (green), I500M (orange) and I500M/F501W (yellow) variants. Each ligand position is represented by the distances from the hydroxyl hydrogen to the N δ of His502 (axis-*x*) and from the benzylic hydrogen (C α) to the flavin N5 (axis-*y*) and colored according to the interaction energy scale.

reported for the oxidation of *p*-methoxybenzyl alcohol, in which the hydrogen in the pro-*R* position is selectively abstracted by the flavin.²¹ To further investigate the stereoselectivity in secondary (\pm)-1-(*p*-methoxyphenyl)-ethanol oxidation, the energy profiles of the diffusion of both enantiomers, from the solvent to the active site of AAO and the I500A, I500M, F501A and I500M/F501W variants, were analyzed in PELE simulations.

According to these simulations, the *S*-enantiomer is capable of reaching the active site of the selected variants with good binding energies and adequate distances from both the

FAD and the catalytic His502 for hydroxyl oxidation (Fig. 8, left). This means that the oxidation of the *S*-enantiomer will be more favorable for these variants than for the native enzyme, being consistent with the experimental kinetic parameters (Table 1). Moreover, the catalytic efficiency (k_{cat}/K_m) values – which cluster into three main groups: native AAO and F501A (low efficiency), I500A (intermediate efficiency) and I500M and I500M/F501W (high efficiency) (Table 1) – qualitatively correlate with the catalytic population observed for the *S*-enantiomer (Fig. 8, left).

In contrast to the above calculations for the *S*-enantiomer, when simulations were performed with (*R*)-1-(*p*-methoxyphenyl)-ethanol in the native enzyme and variants (Fig. 8, right), the alcohol did not reach the active site at distances that would allow the simultaneous abstraction of the proton from the alcohol and the hydride transfer (which in this case should be from the *S* position) in agreement with the experimental data (Fig. 5A).

Conclusions

The ability to oxidize secondary alcohols has been introduced in AAO by site-directed mutagenesis guided by computational simulations with the adaptive PELE software. The I500M/F501W double variant appears as a biocatalyst of biotechnological interest since it produces enantiomerically-enriched secondary alcohols (up to ee >99%), only at the expense of molecular oxygen, during the kinetic resolution of racemic mixtures. According to the *S* stereo-preference of AAO, the *R*-isomers remain unreacted, together with the ketone formed during oxidation of the *S*-isomers (Scheme 1). This stereoselectivity, which agrees with the binding energy at catalytically-relevant positions (near the active-site catalytic base and the flavin cofactor) in the computational diffusion of the *R*- and *S*-isomers, can be exploited for deracemization of chiral secondary alcohols. In summary, we show how computational simulations can guide protein engineering to switch the oxidase preference on primary benzyl alcohols towards enantioselective oxidation of secondary alcohols. Analysis of the resulting variants reveals that the mutations introduced facilitate the entrance and accommodation of bulkier secondary alcohols at the active site of AAO.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by the INDOX (KBBE-2013-7-613549) EU project and by the BIO2017-86559-R (GenoBioref), CTQ2016-79138-R and BIO2016-79106-R projects of the Spanish Ministry of Economy, Industry and Competitiveness, cofinanced by FEDER funds. Pedro Merino (University of Zaragoza, Spain) is acknowledged for his suggestions on chiral HPLC analyses.



References

- R. N. Patel, *Coord. Chem. Rev.*, 2008, **252**, 659–701.
- S. Servi, D. Tessaro and G. Pedrocchi-Fantoni, *Coord. Chem. Rev.*, 2008, **252**, 715–726.
- T. Matsuda, R. Yamanaka and K. Nakamura, *Tetrahedron: Asymmetry*, 2009, **20**, 513–557.
- W. Kroutil, H. Mang, K. Edegger and K. Faber, *Adv. Synth. Catal.*, 2004, **346**, 125–142.
- R. N. Patel, *Biomolecules*, 2013, **3**, 741–777.
- N. J. Turner, *Chem. Rev.*, 2011, **111**, 4073–4087.
- J. Liu, S. Wu and Z. Li, *Curr. Opin. Chem. Biol.*, 2018, **43**, 77–86.
- M. M. Musa, K. I. Ziegelmann-Fjeld, C. Vieile, J. G. Zeikus and R. S. Phillips, *J. Org. Chem.*, 2007, **72**, 30–34.
- C. V. Voss, C. C. Gruber and W. Kroutil, *Angew. Chem., Int. Ed.*, 2008, **47**, 741–745.
- Y. Kawagoshi and M. Fujita, *World J. Microbiol. Biotechnol.*, 1997, **13**, 273–277.
- S. Dieth, D. Tritsch and J. F. Biellmann, *Tetrahedron Lett.*, 1995, **36**, 2243–2246.
- W. Adam, M. Lazarus, B. Boss, C. R. Saha-Möller, H. U. Humpf and P. Schreier, *J. Org. Chem.*, 1997, **62**, 7841–7843.
- E. W. van Hellemond, L. Vermote, W. Koolen, T. Sonke, E. Zandvoort, D. P. Heuts, D. B. Janssen and M. W. Fraaije, *Adv. Synth. Catal.*, 2009, **351**, 1523–1530.
- F. Escalettes and N. J. Turner, *ChemBioChem*, 2008, **9**, 857–860.
- W. P. Dijkman, C. Binda, M. W. Fraaije and A. Mattevi, *ACS Catal.*, 2015, **5**, 1833–1839.
- J. Carro, P. Ferreira, L. Rodríguez, A. Prieto, A. Serrano, B. Balcells, A. Ardá, J. Jiménez-Barbero, A. Gutiérrez, R. Ullrich, M. Hofrichter and A. T. Martínez, *FEBS J.*, 2015, **282**, 3218–3229.
- J. Carro, E. Fernández-Fueyo, M. Alcalde, A. T. Martínez, P. Ferreira, R. Ullrich and M. Hofrichter, *Spain Pat.*, P201730805, 2017.
- F. Guillén, A. T. Martínez and M. J. Martínez, *Eur. J. Biochem.*, 1992, **209**, 603–611.
- P. Ferreira, A. Hernández-Ortega, B. Herguedas, A. T. Martínez and M. Medina, *J. Biol. Chem.*, 2009, **284**, 24840–24847.
- A. Hernández-Ortega, K. Borrelli, P. Ferreira, M. Medina, A. T. Martínez and V. Guallar, *Biochem. J.*, 2011, **436**, 341–350.
- A. Hernández-Ortega, P. Ferreira, P. Merino, M. Medina, V. Guallar and A. T. Martínez, *ChemBioChem*, 2012, **13**, 427–435.
- F. J. Ruiz-Dueñas, P. Ferreira, M. J. Martínez and A. T. Martínez, *Protein Expression Purif.*, 2006, **45**, 191–199.
- A. Aliverti, B. Curti and M. A. Vanoni, *Methods Mol. Biol.*, 1999, **131**, 9–23.
- R. M. C. Dawson, D. C. Elliot, W. H. Elliot and K. M. Jones, *Data for biochemical research*, Oxford Science Publications, Oxford, 1986.
- P. Ferreira, M. Medina, F. Guillén, M. J. Martínez, W. J. H. van Berkel and A. T. Martínez, *Biochem. J.*, 2005, **389**, 731–738.
- S. A. Rounds, F. D. Wilde and G. F. Ritz, *U. S. Geological Survey Techniques of Water Resources Investigations*, book 9, chap. A6, sec. 6.2 (http://water.usgs.gov/owq/FieldManual/Chapter6/6.2_v3.0.pdf), 2013.
- K. Faber and W. Kroutil, <http://biocatalysis.uni-graz.at/enantio/DataFiles/Selectivity-Help.pdf>, 2012.
- D. Lecina, J. F. Gilabert and V. Guallar, *Sci. Rep.*, 2017, **7**, 8466.
- A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin and I. Bahar, *Biophys. J.*, 2001, **80**, 505–515.
- J. F. Gilabert, D. Lecina, J. Estrada and V. Guallar, in *Biomolecular Simulations in Drug Discovery*, ed. F. L. Gervasio and V. Spiwok, Wiley, 2018, pp. 87–96.
- J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio and R. M. Levy, *J. Comput. Chem.*, 2005, **26**, 1752–1780.
- R. B. Murphy, D. M. Philipp and R. A. Friesner, *J. Comput. Chem.*, 2000, **21**, 1442–1457.
- G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, *J. Comput.-Aided Mol. Des.*, 2013, **27**, 221–234.
- A. D. Bochevarov, E. Harder, T. F. Hughes, J. R. Greenwood, D. A. Braden, D. M. Philipp, D. Rinaldo, M. D. Halls, J. Zhang and R. A. Friesner, *Int. J. Quantum Chem.*, 2013, **113**, 2110–2142.
- I. S. Fernández, F. J. Ruiz-Dueñas, E. Santillana, P. Ferreira, M. J. Martínez, A. T. Martínez and A. Romero, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2009, **65**, 1196–1205.
- J. Carro, P. Amengual-Rigo, F. Sancho, M. Medina, V. Guallar, P. Ferreira and A. T. Martínez, *Sci. Rep.*, 2018, **8**, 8121.
- J. Carro, M. Martínez-Júlvez, M. Medina, A. T. Martínez and P. Ferreira, *Phys. Chem. Chem. Phys.*, 2017, **19**, 28666–28675.
- A. Hernández-Ortega, F. Lucas, P. Ferreira, M. Medina, V. Guallar and A. T. Martínez, *J. Biol. Chem.*, 2011, **286**, 41105–41114.
- J. Viña-González, D. González-Pérez, P. Ferreira, A. T. Martínez and M. Alcalde, *Appl. Environ. Microbiol.*, 2015, **81**, 6451–6462.
- J. Viña-González, D. Jiménez-Lalana, A. Serrano, A. T. Martínez and M. Alcalde, *Abs. XL Congr. SEBBM*, Barcelona, 2017, vol. P10-20, p. 90.
- R. J. Fox and M. D. Clay, *Trends Biotechnol.*, 2009, **27**, 137–140.



Supplementary information

Switching the substrate preference of fungal aryl-alcohol oxidase: Towards stereoselective oxidation of secondary benzyl alcohols

Ana Serrano,^{a#} Ferran Sancho,^{b#} Javier Viña-González,^c Juan Carro,^a Miguel Alcalde^c, Victor Guallar,^{b,d} and Angel T. Martínez^{a*}

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

^b Barcelona Supercomputing Center, Jordi Girona 31, E-08034 Barcelona, Spain

^c Department of Biocatalysis, Institute of Catalysis, CSIC, Cantoblanco E-28049, Madrid, Spain

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

[#] These two authors equally contributed to the work

This Supplementary information includes **Tables S1-S3**, **Figs. S1-S4**, and supplementary **references**

Table S1. Spectroscopic properties of AAO and the nine variants analyzed, in 50 mM phosphate, pH 6.0.

	$\lambda^{\text{band I}}$ (nm)	$\lambda^{\text{band II}}$ (nm)	$\epsilon^{\text{band I}}$ (M ⁻¹ cm ⁻¹) ^a
AAO	386	463	11050 ¹
Y92F	386	463	10044 ²
I500A	386	457	9925
I500M	384	458	9609
I500W	388	461	9668
F501A	387	462	10389 ³
F501W	387	462	9944 ³
L315A/I500A	385	458	9904
I391A/I500A	386	458	10089
I500M/F501W	386	460	9290

^aTaken from literature¹⁻³ or estimated here.

Table S2. Chromatographic conditions for secondary alcohol resolution by chiral HPLC.

	<i>n</i> -Hexane/ isopropanol	<i>R</i> (min)	<i>S</i> (min)	Standard (RT, min)
1-(<i>p</i> -Methoxyphenyl)-ethanol	98:2	18.8	20.0	2-phenyl-2-propanol (9.1)
1-Phenyl-1-propanol	98:2	10.1	11.6	2-phenyl-2-propanol (9.1)
1-Phenyl-2-methylpropanol	99:1	12.0	12.5	4-methoxythioanisole (8.8)

Table S3. Oxidation rate (k_{obs}), conversion yield, and (*R*)-1-(*p*-methoxyphenyl)-ethanol *ee* in 24-h reactions of (\pm)-1-(*p*-methoxyphenyl)-ethanol (5 mM) with AAO and nine variants (5 μ M) in 50 mM phosphate, pH 6.0, at 25°C.

	k_{obs} (min ⁻¹)	Conversion (%) ^a	<i>ee</i> (%)
AAO	3.8	34	51
Y92F	1.8	20	25
I500A	57.8	50	100
I500M	-	50	100
I500W	0	0	0
F501A	7.5	46	87
F501W	1.9	7	15
L315A/I500A	4.1	37	59
I391A/I500A	0.6	8	9
I500M/F501W	-	50	100

^aReferred to the racemic mixture

Supplementary references

1. F. J. Ruiz-Dueñas, P. Ferreira, M. J. Martínez and A. T. Martínez, *Protein Express. Purif.*, 2006, **45**, 191-199.
2. P. Ferreira, A. Hernández-Ortega, K. Borrelli, F. Lucas, B. Herguedas, V. Guallar, A. T. Martínez and M. Medina, *FEBS J.*, 2015, **282**, 3091-3106.
3. A. Hernández-Ortega, F. Lucas, P. Ferreira, M. Medina, V. Guallar and A. T. Martínez, *J. Biol. Chem.*, 2011, **286**, 41105-41114.

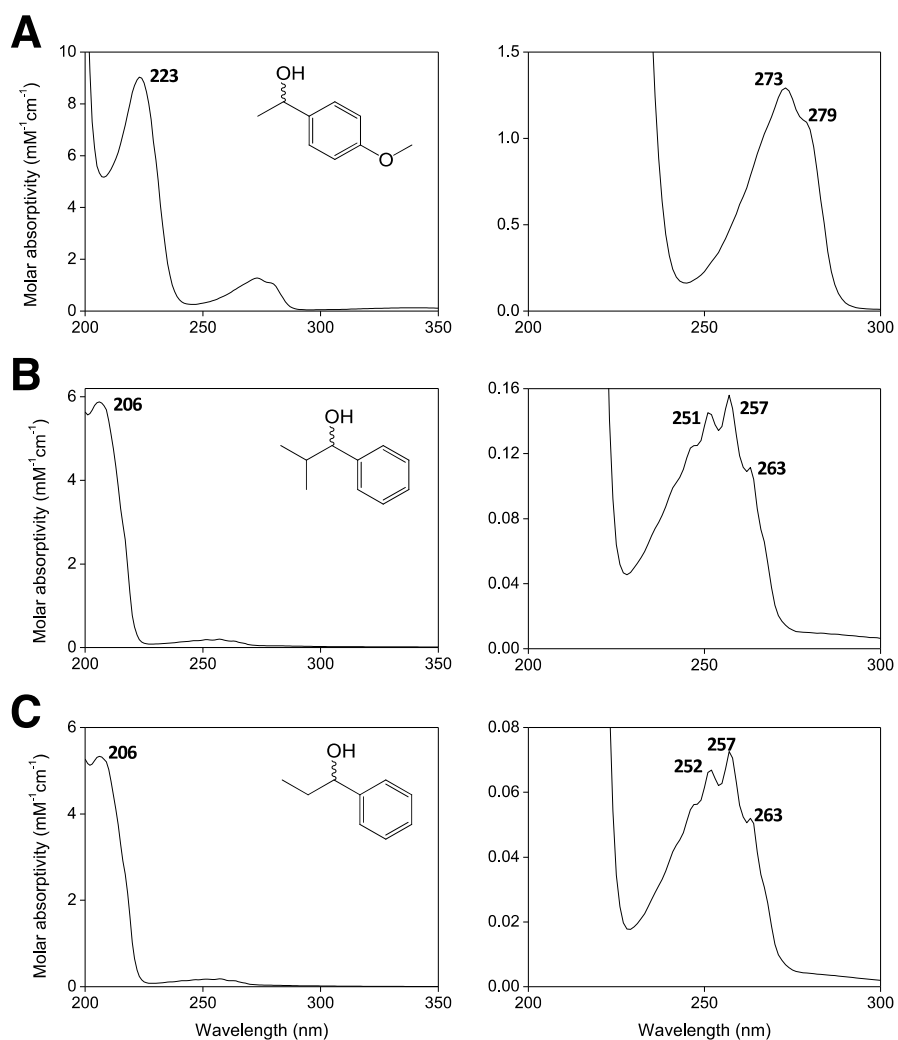


Fig. S1. Molar absorptivity spectra of (\pm) 1-(*p*-methoxyphenyl)-ethanol (**A**), (\pm) 1-phenyl-1-propanol (**B**) and (\pm) 1-phenyl-2-methylpropanol (**C**) in the ultraviolet region (*left*) and enlarged 250-280 nm peak (*right*).

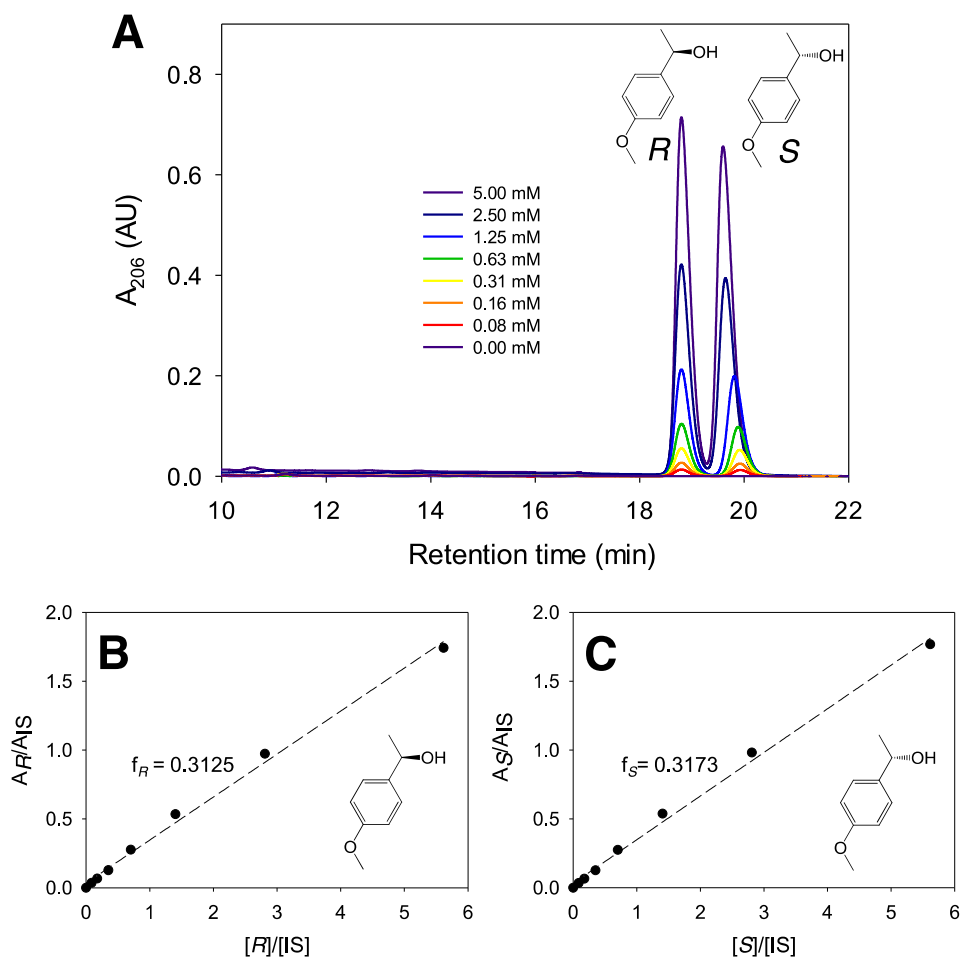


Fig. S2. Separation of different concentrations of (\pm)-1-(*p*-methoxyphenyl)-ethanol by chiral HPLC (A) and calibration curves of *R* (B) and *S* (C) enantiomers yielding response factors (f_R and f_S) using an internal standard (IS), 2-phenyl-2-propanol (not shown).

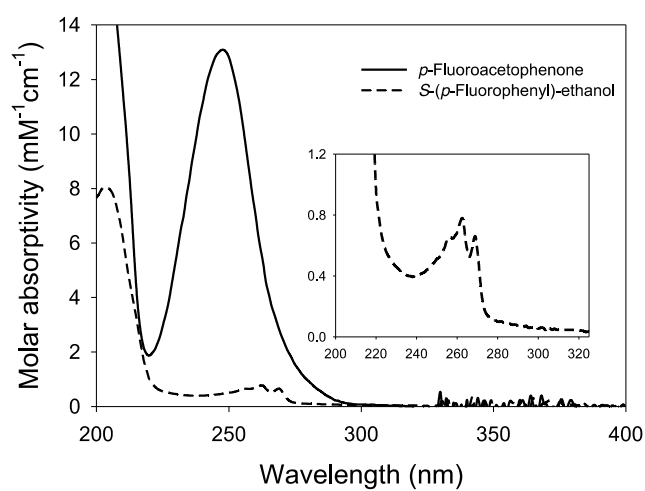


Fig. S3. Molar absorptivity spectra of (*S*)-1-(*p*-fluorophenyl)-ethanol (*dashed line*) and *p*-fluoroacetophenone (*solid line*). The inset shows an enlargement of the alcohol spectrum.

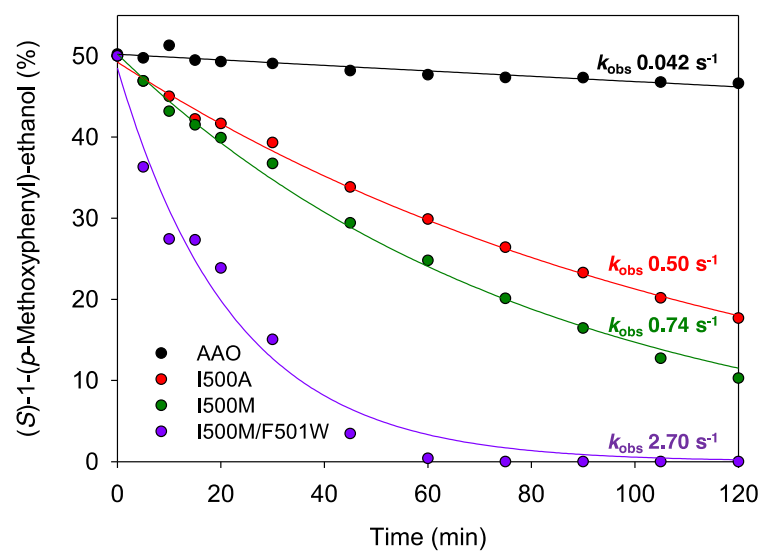


Fig. S4. Oxidation rates of (S)-1-(p-methoxyphenyl)-ethanol by AAO and three selected variants during 2-h incubation. Reactions between alcohol (2.5 mM, racemic mixture) and enzyme (2.5 μ M) were performed in 50 mM phosphate, pH 6.0, at 25°C.

1.3. Article III

Structure-Guided Evolution of Aryl Alcohol Oxidase from *Pleurotus eryngii* for the Selective Oxidation of Secondary Benzyl Alcohols

Javier Viña-Gonzaleza,^a Diego Jimenez-Lalana,^a Ferran Sancho,^b Ana Serrano,^c Angel T. Martinez,^c Victor Guallar,^{b,d} and Miguel Alcalde^{a*}

^a Department of Biocatalysis, Institute of Catalysis, CSIC, Cantoblanco, 28049 Madrid, Spain

Fax: (+31)-91 5854760; phone: (+34)-91 5854806

E-mail: malcalde@icp.csic.es

^b Barcelona Supercomputing Center, Jordi Girona 31, 08034 Barcelona, Spain

^c Biological Research Center, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain

^d Icrea, Passeig Lluís Companys 23, 08010 Barcelona, Spain

Structure-Guided Evolution of Aryl Alcohol Oxidase from *Pleurotus eryngii* for the Selective Oxidation of Secondary Benzyl Alcohols

Javier Viña-Gonzalez,^a Diego Jimenez-Lalana,^a Ferran Sancho,^b Ana Serrano,^c Angel T. Martinez,^c Victor Guallar,^{b, d} and Miguel Alcalde^{a,*}

^a Department of Biocatalysis, Institute of Catalysis, CSIC, Cantoblanco, 28049 Madrid, Spain
Fax: (+31)-91 5854760; phone: (+34)-91 5854806
E-mail: malcalde@icp.csic.es


^b Barcelona Supercomputing Center, Jordi Girona 31, 08034 Barcelona, Spain

^c Biological Research Center, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain

^d ICREA, Passeig Lluís Companys 23, 08010 Barcelona, Spain

Manuscript received: January 29, 2019; Revised manuscript received: March 7, 2019;

Version of record online: April 4, 2019

 Supporting information for this article is available on the WWW under <https://doi.org/10.1002/adsc.201900134>

Abstract: Aryl alcohol oxidase (AAO) is a fungal flavoenzyme capable of oxidizing aromatic primary alcohols into their correspondent aldehydes through a stereoselective hydride abstraction. Unfortunately, this enzyme does not act on secondary benzyl alcohols in racemic mixtures due to the strict control of substrate diffusion and positioning at the active site restricted to primary benzyl alcohols. Here we describe the engineering of AAO from *Pleurotus eryngii* to oxidize chiral benzyl alcohols with high enantioselectivity. The secondary benzyl alcohol oxidase was remodeled at the active site through four cycles of structure-guided evolution, including a final step of *in vivo* site-directed recombination to address the positive epistatic interactions between mutations. The final variant, with five substitutions and a renovated active site, was characterized at biochemical and computational level. The mutational sculpting helped position the bulkier (S)-1-(*p*-methoxyphenyl)-ethanol, improving the mutant's catalytic efficiency by three orders of magnitude relative to the native enzyme while showing a high enantioselectivity (ee >99%). As a promising candidate for racemic resolution, this evolved secondary benzyl alcohol oxidase maintained its natural stereoselective mechanism while displaying activity on several secondary benzyl alcohols.

Keywords: aryl alcohol oxidase; secondary benzyl alcohols; *Saccharomyces cerevisiae*; directed evolution

Introduction

With global annual sales of \$ 1 trillion, an increasingly important challenge in drug development that the pharmaceutical sector must overcome is that posed by chiral chemistry.^[1] Indeed, enantiopure building blocks are in strong demand to produce drugs with particular biological activities, while they are also paramount for the production of fine chemicals.^[2] Biocatalysis has found an important niche in the field of chiral technology, with enzymatic and whole-cell biotransformations offering stereo-, regio- and chemio-selectivity under mild reaction conditions. Classically, two fundamental approaches are followed in the industrial production of chiral molecules: enantioselective synthesis and racemic mixture separation. While asym-

metrical synthesis involves complex and expensive processes,^[3] dynamic kinetic resolution (DKR) is currently one of the most efficient sources of chiral molecules, in which separation is coupled to the *in situ* re-racemization of one of the enantiomers. Another recently described transformative source of enantiomers is cyclic de-racemization. Based on cyclic oxidation-reduction sequences, after the selective oxidation of one of the enantiomers, the achiral intermediate (ketone or imine) is non-selectively reduced to the racemic initial material. After a certain number of cycles, a theoretical 100% yield of the non-reactive enantiomer can be accumulated.^[4] The core step in racemic separation, kinetic resolution, has been achieved for secondary alcohols using either whole-cell systems^[5] or isolated enzymes, such as alcohol

dehydrogenases,^[6] lipases^[7] and particularly, oxidases.^[8] This latter solution is the simplest means to prepare optically pure secondary benzyl alcohols, given their trivial requirements (only needing O₂ from the air) and high enantioselectivity.^[8a,9]

Among these enzymes, aryl alcohol oxidase (AAO, EC 1.1.3.7) is a promising candidate for the enantioselective oxidation of chiral benzyl alcohols. AAO is a flavoenzyme belonging to the GMC (glucose-methanol-choline) superfamily of oxidoreductases and it is naturally secreted as a part of the fungal enzymatic consortium involved in lignin degradation.^[10] Accepting a variety of aromatic alcohols as substrates, the activity of AAO is initially divided into two half-reactions. The first reductive half-reaction involves highly enantioselective hydride transfer from the alcohol's α -C to the FAD co-factor. This process yields the corresponding aldehyde, such that the FAD can then be reoxidized by O₂, releasing H₂O₂ as a by-product of the second oxidative half of the reaction.^[11] Unfortunately, the active site of AAO is buried under a hydrophobic constriction formed by residues Tyr92, Phe397 and Phe501. As a result, substrates bulkier than primary aromatic alcohols cannot easily be accommodated, reducing the enzyme's activity on chiral molecules to a residual trace. Until recently the failure to functionally express AAO in an appropriate heterologous host had prevented its directed evolution. However, fusing the enzyme to a chimeric signal prepro-leader has enabled this protein to be successfully evolved for secretion by yeast.^[12]

Taking advantage of this expression system, here we have combined different laboratory evolution strategies to unlock the activity of AAO on secondary aromatic alcohols. We first carried out a carefully structure-guided campaign of evolution using chiral 1-(*p*-methoxyphenyl)-ethanol as the substrate, thereby generating a palette of secondary benzyl alcohol oxidase mutants. Employing *in vivo* site-directed recombination approaches, mutations were curated by comparing them with their correspondent parental reversions. The differential oxidation of secondary benzyl alcohols by the final benzyl alcohol oxidase variant was characterized, while the rationale behind these changes was analyzed computationally at the atomic level.

Results and Discussion

Laboratory Evolution

First Generation: Unlocking Activity for Secondary Aromatic Alcohols


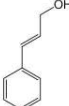
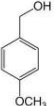

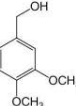
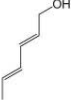
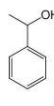
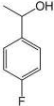

The departure point of this study was a secretion mutant of the AAO from *Pleurotus eryngii* named FX9. This mutant is the product of several rounds of directed evolution aimed to promote functional ex-

pression in *Saccharomyces cerevisiae*.^[12] In this FX9 variant, the AAO is fused to a chimeric prepro-leader (pre α -factor-proKiller) that enhanced secretion by introducing the F[3 α]S, N[25proK]D, T[50proK]A and F[52proK]L mutations into the pre- α and pro-Killer leaders. In addition, two substitutions were included in the mature AAO: L170M in an α -helix situated at the protein surface, and the consensus, ancestral mutation H91N in the FAD attachment loop. These latter mutations enhanced stability and improved production by *S. cerevisiae* to 4.5 mg/L and by *Pichia pastoris* in bioreactor to 25.5 mg/L, while conserving the general biochemical features of the native AAO.^[12]

As a substrate for the screening assay we chose 1-(*p*-methoxyphenyl)-ethanol, a chiral molecule with similar structure to the natural *p*-methoxybenzyl alcohol substrate, the oxidation of which by AAO can be rapidly assessed in a coupled Amplex Red/HRP assay (see Experimental Section for details). As the activity of AAO on secondary alcohols is irrelevant, no response was detected with 1-(*p*-methoxyphenyl)-ethanol when the parental FX9 was screened in microtiter fermentations of the supernatant (i.e. cultures in 96 well plates). As indicated previously, AAO's failure to oxidize secondary aromatic alcohols is due to steric perturbation of the residues forming the catalytic cavity when trying to accommodate bulkier, chiral molecules. Specifically, Phe501 is thought to be a steric liability at the active site, particularly given the very weak but detectable activity on 1-(*p*-methoxyphenyl)-ethanol of a F501A variant expressed in *E. coli* after *in vitro* refolding.^[21] A more recent rational attempt to achieve secondary alcohol oxidation based on PELE (Protein Energy Landscape Exploration) suggested Ile500 was another possible obstacle for ligand diffusion. Indeed, the I500A substitution conferred better transit of the substrate to the active site due to channel broadening.^[22]

To ensure that the activity on secondary benzyl alcohol could be measured at the start of the laboratory evolution campaign, we first prepared three mutants from the FX9 secretion variant: I500A, F501A and I500A-F501A (Figure S2A, B). When isolated from yeast supernatants these variants did not appear to act on 1-(*p*-methoxyphenyl)-ethanol. Hence, we performed combinatorial saturation mutagenesis of the Ile500-Phe501 residues and we found several clones with activity on 1-(*p*-methoxyphenyl)-ethanol, which were scaled up to a 100 mL flask to estimate the overall improvement in activity. Of these, the I500Q-F501W, I500L-F501I, I500M-F501V and I500M-F501W mutants presented a 5-, 15-, 30- and 160-fold enhancement in activity relative to the parental FX9, respectively. The activity of the I500M-F501W mutant (named 15G12) was further tested against a panel of primary and secondary alcohols (Table 1). Its specific activity on primary alcohols was dramatically reduced

Table 1. Specific activity of AAO variants with primary and secondary alcohols.

	Benzyl alcohol	Cinnamyl alcohol	<i>p</i> -Methoxybenzyl alcohol	<i>p</i> -Chlorobenzyl alcohol	3,4-Dimethoxybenzyl alcohol	2,4-Hexadien-1-ol	1-Phenylethanol	<i>p</i> -Fluoro- α -methylbenzyl alcohol	1-(<i>p</i> -Methoxyphenyl)-ethanol
									
FX9 (U/mg)	15.6	41.1	46.6	26.6	22.2	43.4	n.d.	n.d.	2.2x10 ⁻³
15G12 (U/mg)	0.6	18.4	0.3	0.6	0.03	13	6x10 ⁻³	6x10 ⁻³	0.3

Specific activities were estimated in 100 mM phosphate buffer pH 6.0 containing 5 mM of each alcohol. Each reaction was performed by triplicate and substrate conversion was followed by measuring the absorption at 563 nm ($\epsilon_{563} = 56000 \text{ M}^{-1} \text{ cm}^{-1}$) using the HRP/Amplex red coupled assay as described in the experimental section.

depending on the chemical nature of the molecule, with practically no activity on *p*-methoxybenzyl alcohol (c.a. 0.6% of that of the parental FX9 variant). By contrast, the activity of the 15G12 variant on secondary aromatic alcohols rose from undetectable levels, to weak yet evident activity on 1-phenylethanol and 4-fluoro- α -methylbenzyl alcohol, and up to 0.3 U/mg on 1-(*p*-methoxyphenyl)-ethanol (Table 1). Given that hydride abstraction of *p*-methoxybenzyl alcohol to the flavin of AAO occurs exclusively from the pro-R position, in deracemization reactions of secondary alcohols hydrogen abstraction should produce S enantioselective oxidation of the alcohol to the corresponding ketone.

To confirm this, reactions were performed with the optically pure (R) and (S)-4-fluoro- α -methylbenzyl alcohol enantiomers, and as expected, only activity on the latter was detected.^[23]

Paradoxically, while the initial search for a wider space for secondary aromatic alcohol accommodation focused on the I500A-F501A mutations, the bulky alcohol would appear to be much better oxidized in a narrower catalytic pocket following the substitution of Ile500 and Phe501 by the more expansive Met and Trp, respectively (Figure S2C, D). Thus, these two mutations at the active site reposition the secondary alcohol, favouring catalysis, as confirmed by computational analysis (see below).

Second and Third Generations: Searching for new Beneficial Mutations

After unlocking the activity on secondary benzyl alcohols, different protein segments of the 15G12 mutant were targeted for random mutagenesis and

DNA recombination by MORPHING, with a view to further optimize its activity on secondary alcohols. This focused structural evolution tool allows the protein to be divided into defined mutagenic areas, each of which can be interrogated in conjunction with the *in vivo* recombination of the different DNA fragments in *S. cerevisiae*.^[14] This approach has already been successfully applied during the evolution of AAO towards functional expression in yeast^[12] and on this occasion, the design involved the study of three protein blocks in independent libraries (Figure 1A, B). The first of these was the MA block at the N-terminus (Leu48-Thr100), which is associated with the access channel and the FAD-binding domain, and that contains important determinants like Pro79, Asn91, Tyr92 and Val90.

The second MB block (Leu310-Ile417) covers the wall of the catalytic pocket and it contains the aromatic Phe397, a residue implicated in substrate positioning and product release. The third MC block (Glu490-Gln566) is situated in the C-terminal region, and it contains the catalytic His502 and several amino acids related to substrate positioning (Met500, Trp501 and His546). Together, these three blocks encompass the complete active site and the aromatic bottleneck formed by Tyr92, Phe397 and Trp501 (Figure 1C).^[16,23,24]

We carried out the three independent libraries of MORPHING; besides, we prepared a conventional error-prone PCR (ep-PCR) mutagenic library that targeted the whole AAO gene (in total over 4,000 clones were screened). From the pool of libraries, seven mutants with stronger activity on the secondary 1-(*p*-methoxyphenyl)-ethanol were selected for further analysis, ranging from a 1.4 to 2.2-fold enhancement

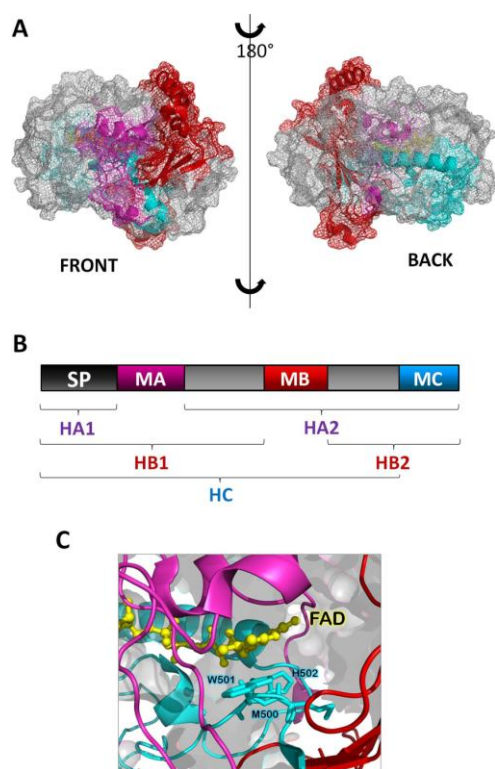


Figure 1. MORPHING fragments for focused evolution. (A) Front and back views of AAO, with the MA, MB and MC MORPHING blocks shown in purple, red and blue, respectively. (B) The dark grey box corresponds to the signal peptide, and the three mutagenic fragments considered for MORPHING are shown as purple, red and blue boxes. For the first MORPHING library, mutagenic block MA was *in vivo* assembled with high-fidelity fragments HA1 and HA2. For the second library, mutagenic fragment MB was recombined from high-fidelity fragments HB1 and HB2. The third library was constructed with the assembly of mutagenic block MC and high-fidelity fragment HC. (C) The catalytic pocket of AAO with the contribution of the MORPHING blocks MA, MB and MC (purple, red and blue, respectively). Model prepared with the crystal structure of the AAO from *P. eryngii* (PDB 3FIM).

over 15G12 (Figure 2). The most successful MORPHING corresponded to the MA block, with five of the seven improved AAO variants from this library. Interestingly, all the mutations identified were located within a 14 amino acid span, from Ile76 to Val90. The mutations carried by the best variant (3F10) were I76V and M83I, which in conjunction with the rest of the substitutions in this stretch (A77V, R80C, I76V and V90A) highlight the importance of the access channel in modulating oxidative activity. Indeed, it should be noted that the R80C substitution was also found in the triple 6G3 mutant from the whole gene ep-PCR library,

together with two superficial substitutions: E39G and Q466R (Figure 2). The MORPHING method was also successful in fragment MB, where the 12D12 variant presented two mutations at the surface: F332L and V340A. Applying focused mutagenesis to the area corresponding to the catalytic pocket, the MC block, was ineffective since variants with improvements were not detected.

To further enhance secondary alcohol oxidation, the best variants from each library were submitted to ep-PCR and *in vivo* shuffling in *S. cerevisiae*: 3F10 and 11H2 from the MA library, the 12D2 variant from the MB library, and the 6G3 variant obtained through whole-gene mutagenic amplification. From this third generation, the 3C11 variant was seen to enhance the activity roughly 1.2-fold compared to 3F10 (412-fold relative to the FX9 parental type), retaining the I76V mutation from 3F10 and acquiring the V90A change, also previously detected in the 9F2 variant, as well as incorporating the new Q174R substitution (Figure 2).

Fourth Generation: Mutational Polishing by *in vivo* site-directed Recombination

After careful evaluation of the mutations obtained in the second and third generations (Table 2), we decided to undertake a final round of evolution to assess whether there were any positive epistatic effects among the mutations. We constructed a combinatorial library by *in vivo* site-directed recombination, such that the 10 mutations and their corresponding reversions could be rapidly combined in an one-pot transformation reaction, evaluating the library in order to obtain the optimal combination of substitutions for the oxidation of chiral alcohols (Figure S1). From this ensemble of mutations, the three best variants identified shared the same backbone of substitutions: A77V-R80C-V340A. The third best variant was the 24H3 mutant that carried the A77V-R80C-Q174R-F332L-V340A-Q466R muta-

Table 2. Selected mutations for site-directed recombination.

Mutation	Variant	Library	Secondary motif
GAG E39G _{GGG}	6G3	ep-PCR	Loop
ATT I76V _{GTT}	3F10, 4D8	MA	Loop
GCG A77V _{GTG}	11H2	MA	Loop
CGC R80C _{TGC}	6G3, 7B4	ep-PCR, MA	Loop
ATG M83I _{ATA}	3F10	MA	Loop
GTT V90A _{GCT}	9F2	MA, mutagenic shuffling	Loop
CAA Q174R _{CGA}	3C11	Mutagenic Shuffling	Alpha helix
TTC F332L _{CTC}	12D2	MB	Alpha helix
GTT V340A _{GCT}	12D2	MB	Alpha helix
CAA Q466R _{CGA}	6G3	ep-PCR	Loop

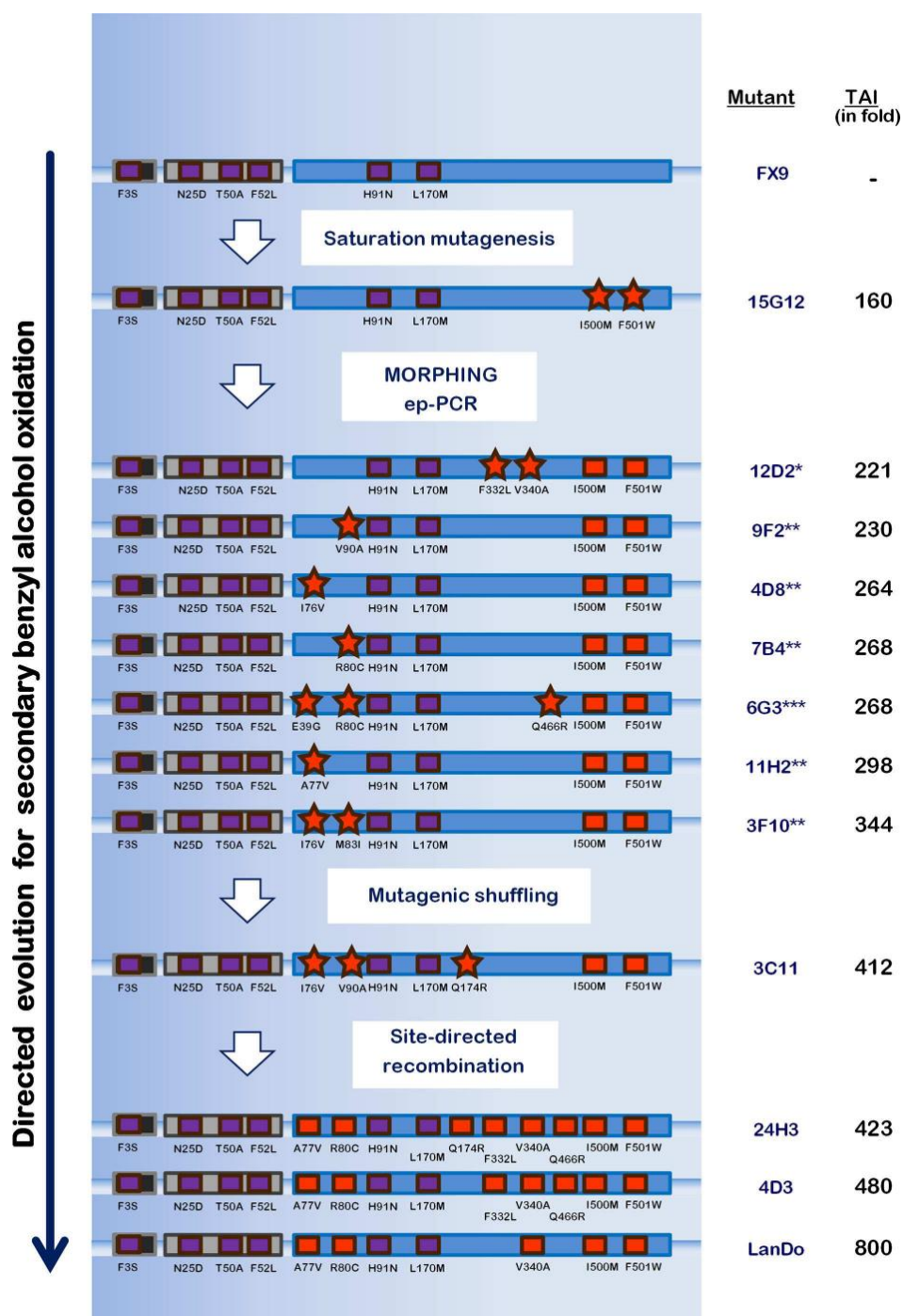


Figure 2. Laboratory evolution of AAO for the oxidation of secondary benzyl alcohols. New mutations are represented as stars and accumulated mutations as squares. The chimeric prepro-leader is depicted in grey and the mature AAO in blue. The TAI (total activity improvement) refers to the fold improvement of AAO activity with 1-(*p*-methoxyphenyl)-ethanol as a substrate and it was estimated relative to the FX9 parental type from *S. cerevisiae* supernatants: *Mutants from the MB library; **Mutants from the MA library; ***Mutant from the ep-PCR library.

tions, displaying activity more than 420-fold better than the FX9 parental type. In the case of 4D3, the only difference from the 24H3 variant was the absence of the Q174R mutation, which translated into a 480-fold increase in activity, highlighting a detrimental effect of Q174R within this mutational context. The further purging of Q466R and F332L gave rise to the LanDo variant that carried the A77V-R80C-V340A mutations in conjunction with I500M-F501W and the 6 secretion mutations of FX9, this variant representing the best performer with a total 800-fold enhancement of activity relative to the parental type (Figures 2, 3).

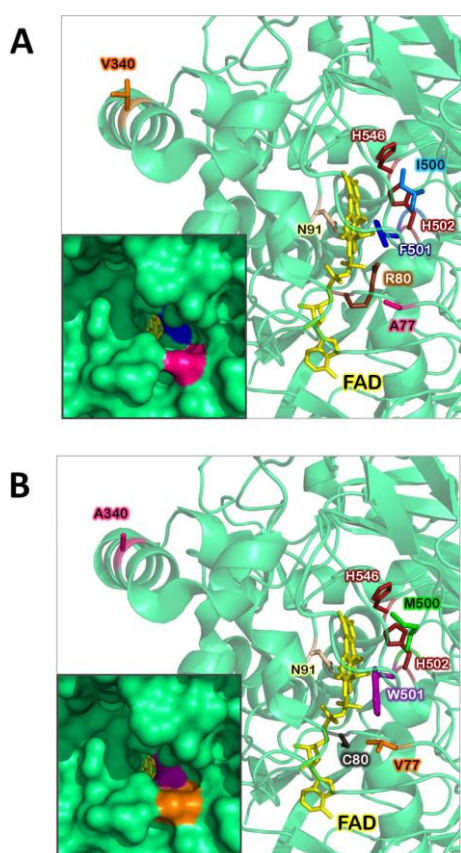


Figure 3. Location of the mutations in the evolved secondary benzyl alcohol oxidase. (A) FX9 parental type and (B) LanDo mutant. The FAD molecule is represented in yellow, the catalytic base His502 and His546 is depicted in red, and the consensus ancestral mutation Asn91 is in light pink. A77V, R80C, V340A, I500M and F501W are represented following a color code (before and after mutation). The mutation L170M at the surface of the enzyme is not present in the fragment represented. The inset shows a detail of the protein surface at the access channel. The model was prepared with a crystal structure of the AAO from *P. eryngii* (PDB 3FIM).

Biochemical Characterization

To characterize the LanDo mutant and the FX9 parental variant, they were produced and purified to homogeneity. To determine the enantioselectivity of the LanDo variant, transformation of the racemic 1-(*p*-methoxyphenyl)-ethanol was followed by chiral-HPLC (Figure 4) and 100% conversion with an enantiomeric

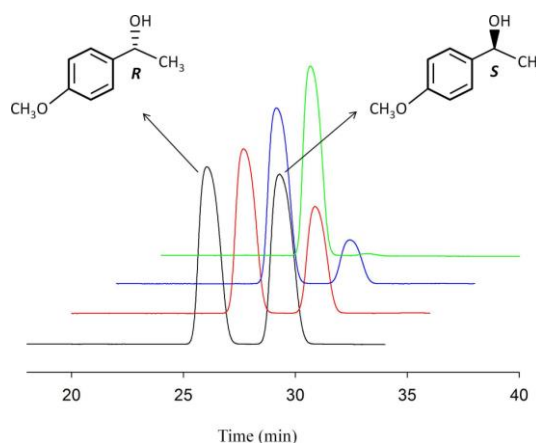


Figure 4. Chiral HPLC analysis. HPLC elution profiles after the reaction of the AAO variant LanDo (1 μ M) with racemic 1-(*p*-methoxyphenyl)-ethanol (2.5 mM). Reactions were performed at room temperature in 100 mM phosphate buffer pH 6.0 with continuous shaking and the aliquots were analyzed by chiral HPLC at different times. The separation of the R and S enantiomers in the negative control is represented in black, whereas the 15, 45, and 90 min reactions are represented in red, blue and green, respectively.

excess (ee) >99% was achieved after a two hour reaction. The configuration of the remaining alcohol in the reaction was confirmed by optical rotation (Figure S3, Table S2), the positive rotation corresponding to the R enantiomer meaning the natural oxidation of the S enantiomer by AAO was maintained after evolution.^[23] Despite the remarkably specific activity for the secondary 1-(*p*-methoxyphenyl)-ethanol (2.9 U/mg), the five new mutations carried by the LanDo variant did not negatively affect its secretion (4.6 mg/L). The activity of LanDo with secondary alcohols was tested against available commercial secondary (aromatic) alcohols that were representative of the structural scope of the AAO. The initial turnover rates of the LanDo variant relative to the wildtype AAO (wtAAO, heterologous expressed in *E. Coli* after *in vitro* refolding) increased 30, 20 and 100-fold times for 1-(*p*-methoxyphenyl)-ethanol, *p*-fluoro- α -methylbenzyl alcohol and 1-phenylethanol, respectively. Indeed, the LanDo variant even showed activity on 1-

phenylpropanol, a substrate not oxidized by wtAAO (Table 3). The kinetic parameters were measured under

Table 3. Initial turnover rates for secondary alcohols.

	<i>p</i> -Fluoro- α -methylbenzyl alcohol	1-Phenylethanol	1-Phenylpropanol
wtAAO	0.35 \pm 0.01	0.10 \pm 0.03	n.m.
LanDo	7.3 \pm 0.1	10 \pm 0.5	1 \pm 0.1

Turnover rates (min^{-1}) were estimated in 100 mM phosphate buffer pH 6.0 containing 5 mM of each secondary alcohol with the exception of 1-phenylpropanol (2.5 mM). Each reaction was performed by triplicate and substrate conversion was followed by measuring the absorption at 563 nm ($\epsilon_{563} = 56000 \text{ M}^{-1} \text{ cm}^{-1}$) using the HRP/Amplex red coupled assay as described in the experimental section.

air-saturated conditions for 1-(*p*-methoxyphenyl)-ethanol, and for the primary alcohols *p*-methoxybenzyl alcohol and 2,4-hexadien-1-ol. LanDo displayed an outstanding increase in the catalytic efficiency for enantioselective oxidation of 1-(*p*-methoxyphenyl)-ethanol, three orders of magnitude. Interestingly, the activity on primary alcohols that was dramatically reduced after inserting the I500M-F501W pair in the first cycle of evolution in 15G12 variant (Table 1, Figure 2), was recovered to a considerable extent for *p*-methoxybenzyl alcohol and the aliphatic 2,4-hexadien-1-ol. This result indicates the beneficial effect that A77V, R80C and V340A exerted on LanDo's overall activity (Table 4).

Computational Analysis

In order to rationalize the effect of the mutations identified, PELE simulations were run for wtAAO and the variants obtained in the different rounds of directed evolution. The oxidation of alcohols by AAO involves a non-synchronous concerted reaction, where both

proton transfer from the hydroxyl group to the catalytic base His502 and hydride abstraction from the benzylic position by the flavin are taking place at the same reaction step.^[11] PELE results were plotted placing both catalytic distances in the X and Y axis, and the interaction energy between the protein and the ligand was represented by colors (Figure 5).

No significant differences were evident between the wtAAO and the secretion mutant FX9 (H91N-L170M) in these plots, consistent with the experimental evidence that this variant does not improve the activity on secondary alcohols but does increase expression and stability. The 15G12 variant included the I500M and F501W mutations on top of the previous ones, accumulating a total of 4 mutations in the mature protein. In this case, the plot shows how the ligand can reach catalytic positions 2.5 Å away from both the FAD and the histidine at the same time, producing better catalytic constants than those of the wtAAO. In addition, a minimum could be seen where the ligand-histidine distance was ~ 2.2 Å, although the interaction energies were much higher and they were therefore less accessible.

The largest improvement came after introducing an additional three mutations in the LanDo variant: A77V, R80C, and V340A. These three substitutions allow the ligand to achieve even smaller catalytic distances, up to ~ 2 Å for the histidine and 2.4 Å for the FAD, with reasonable interaction energies. Considering that closer catalytic distances imply a decrease in energy barriers, this agreed well with the higher kinetic constants. Moreover, the increase in the number of structures with good catalytic distances could reflect the ease with which the ligand can find catalytic positions, explaining the lower K_m values for this variant.

Despite the large number of mutations in LanDo (7 in total, excluding the mutations in the chimeric prepro-leader), there are no major conformational changes in the protein or in the positioning of the ligand. Nevertheless, we did note subtle modifications that were sufficient to improve the catalytic position of

Table 4. Kinetic parameters for AAO variants.

Substrate	Kinetic constants	wtAAO**	LanDo
1-(<i>p</i> -methoxyphenyl)-ethanol	K_m (mM)*	24.9 \pm 1.1	0.65 \pm 0.1
	k_{cat} (s^{-1})	0.18 \pm 0.002	4.9 \pm 0.1
	k_{cat}/K_m ($\text{mM}^{-1} \text{ s}^{-1}$)	0.007	7.5
<i>p</i> -methoxybenzyl alcohol**	K_m (mM)	0.027 \pm 0.004	0.02 \pm 0.003
	k_{cat} (s^{-1})	142 \pm 5	72 \pm 3
	k_{cat}/K_m ($\text{mM}^{-1} \text{ s}^{-1}$)	5233	3600
2,4-hexadien-1-ol**	K_m (mM)	0.094 \pm 0.005	0.095 \pm 0.006
	k_{cat} (s^{-1})	119 \pm 2	40.9 \pm 0.7
	k_{cat}/K_m ($\text{mM}^{-1} \text{ s}^{-1}$)	1271	430.5

AAO kinetic constants were measured in 100 mM phosphate buffer pH 6.0 at 25 °C. All reactions were performed by triplicate.

*Referred to the S enantiomer, as 50% of the racemic mixture. **Calculated for wtAAO as described previously.^[22,26]

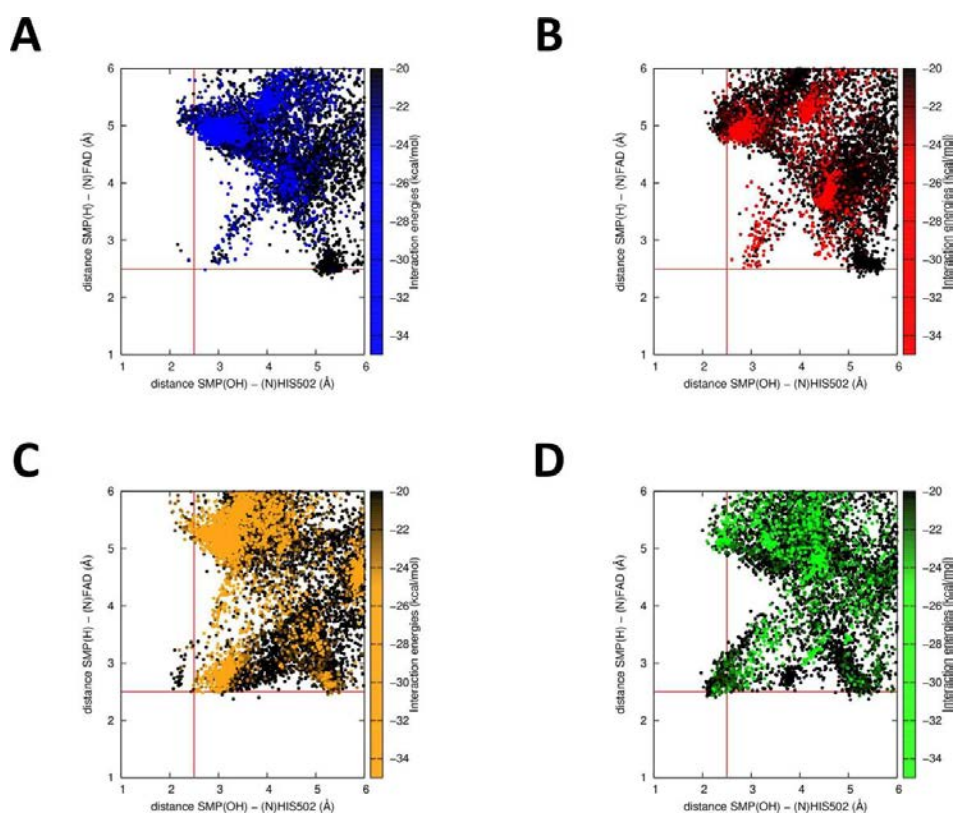


Figure 5. Substrate diffusion computational PELE simulations. Plots represent the PELE simulations relating catalytic distances (X and Y) and interaction energies (color scheme, right Y axis) for different AAO variants: (A) wtAAO; (B) the FX9 secretion mutant; (C) the 15G12 mutant; (D) the final evolved benzyl alcohol oxidase, LanDo mutant.

(S)-1-(*p*-methoxyphenyl)-ethanol. In particular, the R80C mutation was found repeatedly in independent libraries during evolution (Figure 2) and it created an empty space at the top of the FAD cofactor. Consequently, the backbone containing H502 shifts in that direction (Figure 6A). Moreover, Arg80 interacts with the oxygen of the backbone of residue 501, such that this mutation frees Trp501 to form a hydrogen bond with the oxygen backbone of Val77 (Figure 6B). All these subtle adjustments allow the ligand to adopt conformations with better catalytic distances (Figures 5, 6).

Conclusion

Focusing evolution on structural elements made it possible to identify mutations in the catalytic pocket and access channel that allowed an AAO to be designed that acts on different secondary benzyl alcohols. This final secondary benzyl alcohol oxidase variant maintained strong enantioselectivity, providing

a potential catalyst for chiral de-racemization. The complex enzyme-substrate relationships of this enzyme were highlighted by an enhancement of three orders of magnitude in the catalytic efficiency, an effect produced by a combination of bulky substitutions in the catalytic cavity and other unpredicted changes. Paradoxically, a steric problem appeared to be resolved by introducing bulkier residues, something difficult to anticipate from a rational point of view. It is worth noting that the resolution of molecules like 1-phenylpropanol could be of use to obtain moiety precursors of serotonin/norepinephrine reuptake inhibitors like Fluoxetine or Atroxetine.^[6]

The results obtained here also highlight the importance and efficacy of *S. cerevisiae* as a platform for both the functional expression of eukaryotic genes and as a molecular tool-box to generate DNA libraries for directed evolution campaigns. The data presented open new opportunities for the evolution of AAO, which include the oxidation of furfural derivatives for the synthesis of biopolymers or the *in situ* production

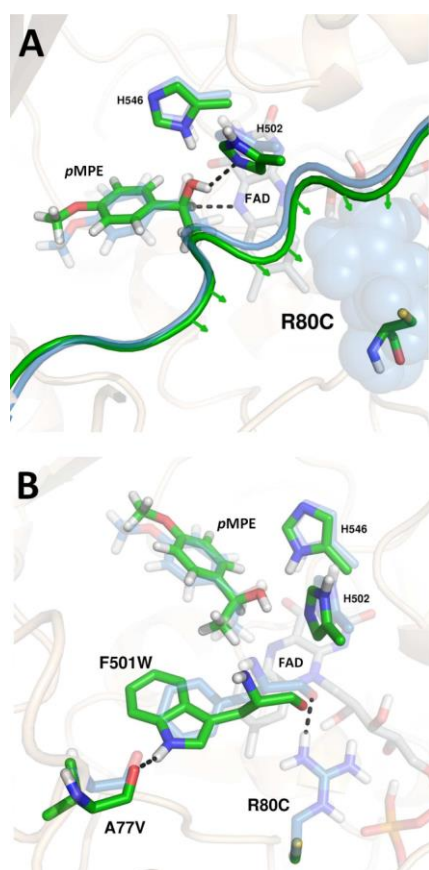


Figure 6. Conformational changes at the catalytic pocket of AAO. (A) Backbone displacement (green) to better position (S)- 1-(p-methoxyphenyl)-ethanol (pMPE). (B) Interruption of the interaction of Arg80 with the backbone of residue 501, and the formation of a hydrogen bond between Val77 and Trp501.

of H_2O_2 in cascade oxyfunctionalization reactions by peroxygenases.^[25]

Experimental Section

Materials

All chemicals were reagent-grade purity. Benzyl alcohol, *p*-chlorobenzyl alcohol, 3,4-dimethoxybenzyl alcohol, *p*-Methoxybenzyl alcohol, 1-phenylethanol, 1-phenylpropanol, 1-(*p*-methoxyphenyl)-ethanol, 4-fluoro- α -methylbenzyl alcohol, (R)-4-fluoro- α -methylbenzyl alcohol, (S)-4-fluoro- α -methylbenzyl alcohol, 2,4-hexadien-1-ol, cinnamyl alcohol, Horseradish peroxidase (HRP), Taq polymerase and the Yeast Transformation Kit were purchased from Sigma (Madrid, Spain). Amplex[®] Red reagent (10-acetyl-3,7-dihydroxyphenoxazine) was obtained from Biogen (Madrid, Spain). Zymoprep Yeast Plasmid Miniprep, Yeast Plasmid Miniprep Kit I and Zymoclean Gel

DNA Recovery Kit were from Zymo Research (Orange, CA, USA). Restriction enzymes BamHI and XhoI were from New England Biolabs (Hertfordshire, UK). I-Proof high-fidelity DNA polymerase was from Biorad (Hercules, CA, USA). The episomal shuttle vector pJRoC30 was from the California Institute of Technology (CALTECH, USA). The protease deficient *S. cerevisiae* strain BJ5465 was from LGC Promochem (Barcelona, Spain). *E. coli* XL2-Blue competent cells were from Stratagene (La Jolla, CA, USA). Primers were acquired from Isogen Life Science (Barcelona, Spain) and are included in Table S1.

Culture Media

Minimal medium SC contained 100 mL 6.7% (w:v) sterile yeast nitrogen base, 100 mL 19.2 g/L sterile yeast synthetic drop-out medium supplement without uracil, 100 mL sterile 20% raffinose (w:v), 700 mL sddH₂O and 1 mL 25 g/L chloramphenicol. YP medium contained 10 g yeast extract, 20 g peptone and ddH₂O to 650 mL whereas YPD medium also contained 20% glucose (w:v). AAO expression medium contained 144 mL YP 1.55 \times , 13.4 mL 1 M KH₂PO₄ pH 6.0 buffer, 22.2 mL 20% galactose (w:v), 0.222 mL 25 g/L chloramphenicol and ddH₂O to 200 mL. Luria Broth (LB) medium contained 10 g sodium chloride, 5 g yeast extract, 10 g peptone, 1 mL 100 mg/mL ampicillin and ddH₂O to 1 L. AAO selective expression medium (SEM) contained 100 mL 6.7% (w:v) sterile yeast nitrogen base, 100 mL 19.2 g/L sterile yeast synthetic drop-out medium supplement without uracil, 100 mL sterile 20% galactose (w:v), 100 mL 1 M KH₂PO₄ pH 6.0, 600 mL sddH₂O and 1 mL 25 g/L chloramphenicol.

Construction of Variants I500A, F501A and I500A-F501A

All the PCR products were cleaned, concentrated, loaded onto a preparative agarose gel (1%, w:v) and purified using the Zymoclean Gel DNA Recovery kit before being cloned into the shuttle vector pJRoC30 under the control of the GAL1 promoter. BamHI and XhoI were used to linearize the plasmid pJRoC30 and to remove the parent gene. FX9 variant was amplified from pJRoC30-FX9^[12b] with two PCR reactions for each mutant containing overhang segments for the whole plasmid to be reassembled in the yeast. PCR reactions were carried out in a final volume of 50 μ L containing 3% DMSO, 0.8 mM dNTPs (0.2 mM each), 0.03 U/ μ L Iproof DNA polymerase, and 0.2 ng/ μ L template. The oligos used for each PCR reaction were: For I500 A, PCR1 (oligo sense RMLN and oligo antisense I500Ar), PCR2 (oligo sense I500Af and oligo antisense RMLC). For F501 A, PCR1 (RMLN and oligo antisense F501Ar), PCR2 (oligo sense F501Af and RMLC). For double mutant I500A-F501 A, PCR1 (RMLN and oligo antisense DM500-1Ar), PCR2 (oligo sense DM500-1Af and RMLC). Amplification reactions were carried out in a thermal cycler Mycycler[™] (BIO-RAD, USA) with the following PCR program: 98 °C for 30 seconds (1 cycle); 98 °C for 10 seconds, 50 °C for 25 seconds and 72 °C for 60 seconds (28 cycles); and 72 °C for ten minutes (1 cycle). After purification, PCR products (400 ng each) were mixed with the linearized pJRoC30 (100 ng; ratio PCR product: vector=4:1) and transformed in

yeast (Yeast transformation kit) for the recombination and *in vivo* cloning. 176 colonies were picked, expressed and screened as described below. Each construct was recovered and its sequence confirmed by DNA sequencing.

Directed Evolution

First generation: combinatorial saturation mutagenesis at positions Ile500 and Phe501: Two PCR reactions were carried out in a final volume of 50 μ L containing 3% DMSO, 0.8 mM dNTPs (0.2 mM each), 0.03 U/ μ L iproof DNA polymerase, and 0.2 ng/ μ L FX9 template and different primers according to the 22-trick protocol.^[13] PCR1 contained 0.25 μ M RMLN and 0.25 μ M mix of reverse primers: 22c1R, 22c2R, 22c3R, 22c4R, 22c5R, 22c6R, 22c7R, 22c8R and 22c9R. PCR2 contained 0.25 μ M RMLC and 0.25 μ M mix of forward primers: 22c1F, 22c2F, 22c3F, 22c4F, 22c5F, 22c6F, 22c7F, 22c8F and 22c9F. Amplification reactions were carried out in a thermal cycler MycyclerTM (BIO-RAD, USA) with the following PCR program: 98 °C for 30 seconds (1 cycle); 98 °C for 10 seconds, 50 °C for 25 seconds and 72 °C for 60 seconds (28 cycles); and 72 °C for ten minutes (1 cycle). After purification, PCR products (400 ng each) were mixed with the linearized pJRoC30 (100 ng; ratio PCR product: vector=4:1) and transformed in yeast for *in vivo* cloning. According to the 22-trick protocol, a library of 3066 individual colonies was screened as described below.

Second generation: MORPHING: The 15G12 variant was used as the parental template for focused random mutagenesis technique MORPHING (Mutagenic Organized Recombination Process by Homologous *In vivo* Grouping).^[14] Three libraries were constructed independently targeting 3 protein blocks: MA, MB, and MC. Additionally, a mutagenic library subjecting the whole AAO fusion was prepared by error prone PCR (ep-PCR). Primers were designed to create homologous overlapping areas of ~50 bp for the whole gene to be reassembled *in vivo* upon transformation in *S. cerevisiae*. i) ep-PCR for MORPHING blocks and whole AAO gene were carried out in a final volume of 50 μ L containing: 90 nM oligo sense (FM1 for MA block, FM2 for MB block, FM3 for MC block and RMLN for whole gene amplification), 90 nM oligo antisense (RM1 for fragment MA, RM2 for fragment MB, RM3 for fragment MC and RMLC for whole gene amplification), 0.3 mM dNTPs (0.075 mM each), 3% DMSO, 0.05 or 0.01 mM MnCl₂, 1.5 mM MgCl₂, 0.05 U/ μ L Taq polymerase DNA, and 0.92 ng/ μ L 15G12 template. The amplification parameters were 95 °C for 2 min (1 cycle); 95 °C for 45 s, 50 °C for 45 s, and 74 °C for 2 min (28 cycles); and 74 °C for 10 min (1 cycle). Concentrations of 0.05 and 0.01 mM MnCl₂ were used for MORPHING and full gene ep-PCR, respectively, to adjust the mutational rate to 1–3 mutations per gene. ii) High-fidelity PCRs for MORPHING were carried out in a final volume of 50 μ L containing 3% DMSO, 0.8 mM dNTPs (0.2 mM each), 0.03 U/ μ L iproof DNA polymerase, 0.25 μ M oligo sense (RMLN for HA1, HB1 and HC fragments, FHF1 for HA2 and FHF2 for HB2), 0.25 μ M oligo antisense (RHF1 for HA1, RHF2 for HB1, RHF3 for MC and RMLC for HA2 and HB2 fragments) and 0.2 ng/ μ L template. High-fidelity PCR was performed using the following parameters: 98 °C for 30 s (1 cycle); 94 °C for 10 s, 48 °C for 30 s, 72 °C for 30 s (30 cycles); and 72 °C for 10 min (1 cycle).

The assembly of the fragments for the different libraries is described in Figure 1B. PCR products were cleaned, concentrated, loaded onto a preparative agarose gel, purified, mixed in equimolar amounts, (200 ng mutagenic fragment and 200 ng non-mutagenic fragment) and transformed with linearized pJRoC30 (200 ng) into chemically competent cells, as described above.

Third generation: mutagenic shuffling: ep-PCR reactions were performed separately with mutants 3F10, 11H2, 12D2 and 6G3. Reaction mixtures were prepared in a final volume of 50 μ L containing DNA template (0.92 ng/ μ L), 90 nM oligo sense RMLN (5'-CCTCTATACTTTAACGTC AAGG-3'), 90 nM reverse primer RMLC (5'-GGGAGGGCGTGAATGTAAGC-3'), 0.3 mM dNTPs (0.075 mM each), 3% (v:v) dimethylsulfoxide (DMSO), 1.5 mM MgCl₂, 0.1 mM MnCl₂ and 0.05 U/ μ L Taq DNA polymerase. PCRs were performed in a thermocycler (Mycycler, Bio-Rad, Hercules, CA, USA) and parameters were: 95 °C for 2 min (1 cycle); 95 °C for 45 s, 50 °C for 45 s, 74 °C for 45 s (28 cycles); and 74 °C for 10 min (1 cycle). The PCR products were mixed with linearized pJRoC30 (at a PCR product/linearized plasmid ratio of 4:1) and transformed into competent *S. cerevisiae* cells to promote *in vivo* DNA shuffling. Transformed cells were plated on SC (synthetic complete) drop-out plates and incubated for 3 days at 30 °C. Colonies containing the whole autonomously replicating vector were picked and subjected to high-throughput screening as described below.

Fourth generation: site-directed recombination: The 3C11 variant was used as template and primers were designed for the 10 selected mutations (E39G, I76V, A77V, R80C, M83I, V90A, Q174R, F332L, V340A, and Q466R) for the *in vivo* site-directed recombination of mutations vs. corresponding reversions at each position of the combinatorial library. A total of 6 PCR reactions were performed. Reactions were carried out in a final volume of 50 μ L containing 3% DMSO, 0.8 mM dNTPs (0.2 mM each), 0.03 U/ μ L iproof DNA polymerase, and 0.2 ng/ μ L 3C11 with 0.25 μ M of the following oligos: PCR1 used oligo sense RMLN and oligo antisense R116 (containing position 39). PCR2 was performed with oligo sense F116 (containing position 39) and oligo antisense R5 (containing positions 76, 77, 80, 83 and 90). PCR3 used oligo sense F5 (containing positions 76, 77, 80, 83 and 90) and oligo antisense R521 (containing position 174). PCR4 used oligo sense F521 (containing position 174) and oligo antisense R9419 (containing positions 332 and 340). PCR5 was performed with oligo sense F9419 (containing positions 332 and 340) and oligo antisense R1397 (containing position 466). PCR6 was performed with oligo sense F1397 (containing position 466) and oligo antisense RMLC, (Figure S1, Table S1). For the *in vivo* assembly of the whole gene, the fragments were designed with overhangs of around 40 bp between them. Amplification reactions were carried out in a thermal cycler and the PCR program was: 98 °C for 30 seconds (1 cycle); 98 °C for 10 seconds, 50 °C for 25 seconds and 72 °C for 30 seconds (28 cycles); and 72 °C for ten minutes (1 cycle). After purification, PCR products (400 ng each) were mixed with the linearized pJRoC30 (100 ng; ratio PCR product: vector=4:1) and transformed in yeast for *in vivo* cloning. A library of 3070 individual colonies was screened as described below.

High-throughput Screening (HTS) Assay

Transformed cells were plated in SC drop-out plates and incubated for 3 days at 30 °C, individual clones were fermented in sterile 96-well plates containing 200 μ L of SEM medium. Plates were sealed and incubated at 30 °C, 225 rpm and 80% relative humidity in a humidity shaker (Minitron-INFORS, Biogen, Spain) for 72 hours. Aliquots of 20 μ L of yeast supernatants were transferred to a 96-well plate using a robotic station for liquid handling Freedom EVO (Tecan, Männedorf, Switzerland) and 180 μ L of HRP-Amplex Red reagent for secondary alcohol activity detection. The final concentrations in the well were 5 mM 1-(*p*-methoxyphenyl)-ethanol, 70 μ M Amplex Red, 3 μ g/mL HRP in 100 mM phosphate buffer pH 6.0. Reagents were dispensed with Multidrop™ Combi Reagent Dispenser (Thermo Scientific, Massachusetts, USA). The plates were incubated at room temperature and activity with the chiral alcohol was determined as H₂O₂ production coupled to the oxidation of Amplex Red reagent by the HRP and measured at 563 nm ($\epsilon_{\text{resorufin } 563} = 56000 \text{ M}^{-1} \text{ cm}^{-1}$). Reaction mixture with *p*-methoxybenzyl alcohol (1 mM final concentration) was also prepared to determine activity with a primary alcohol. One unit of AAO activity is defined as the amount of enzyme that converts 1 μ mol of alcohol to aldehyde or ketone with the stoichiometric formation of H₂O₂ per min under the reaction conditions. The HTS-assay incorporated two consecutive re-screenings to rule out the selection of false positives as described in previous work.^[12a]

Protein Production and Purification

The native AAO, heterologously expressed in *E. coli* and in vitro refolded (wtAAO), was produced and purified as described elsewhere.^[15] Evolved variants were produced in yeast^[12a] and purified by cationic exchange chromatography, anion exchange chromatography and size-exclusion chromatography (ÄKTA purifier, GE Healthcare, WI, US). The crude extract from *S. cerevisiae* cultures was concentrated and dialyzed in 20 mM sodium phosphate/citrate at pH 3.3 (buffer A) by tangential ultrafiltration (Pellicon; Millipore, USA) through a 10-kDa-pore-size membrane (Millipore, USA) by means of a peristaltic pump (Masterflex easy load; Cole-Parmer, USA). The sample was filtered and loaded onto a strong cation-exchange column (HiTrap SP FF; GE Healthcare) pre-equilibrated with buffer A and coupled to the ÄKTA purifier system. The proteins were eluted with a linear gradient of buffer P (20 mM piperazine buffer pH 5.5) + 1 M NaCl in two phases at a flow rate of 1 ml/min: from 0 to 50% during 15 min and from 50 to 100% in 2 min. Fractions with AAO activity were pooled, dialyzed against buffer P, concentrated and loaded onto a high-resolution resin, strong-anion-exchange column (Biosuite MonoQ 10 cm; Waters, USA) pre-equilibrated in buffer P. The proteins were eluted with a linear gradient from 0 to 0.5 M of NaCl in two phases at a flow rate of 1 ml/min: from 0 to 50% during 20 min and from 50 to 100% in 2 min. Fractions with AAO activity were pooled, dialyzed against 20 mM phosphate buffer pH 6.0, concentrated and further purified by a Superose 12 HR 10/30 molecular exclusion column (Amersham Bioscience; Amersham, UK) pre-equilibrated with 150 mM NaCl in 20 mM phosphate buffer pH 6.0 at a flow rate of 0.5 ml/min. The fractions with AAO activity were pooled, dialyzed against

buffer 20 mM phosphate buffer pH 6.0, concentrated and stored.

Biochemical Characterization

Steady-state kinetic constants: Alcohol oxidation kinetics for 1-(*p*-methoxyphenyl)-ethanol, *p*-methoxybenzyl alcohol and 2,4-hexadien-1-ol were measured in 100 mM phosphate buffer pH 6.0 at 25 °C in air-saturated conditions (0.256 mM O₂ concentration). Reactions were performed by triplicate and substrates oxidations were followed by measuring the absorption at 285 nm for *p*-methoxybenzyl alcohol, $\epsilon_{285} = 16,950 \text{ M}^{-1} \text{ cm}^{-1}$ and at 280 nm for 2,4-hexadien-1-ol, $\epsilon_{280} = 30,140 \text{ M}^{-1} \text{ cm}^{-1}$. The oxidation of 1-(*p*-methoxyphenyl)-ethanol was measured indirectly coupled with saturated conditions of HRP and Amplex Red substrate (4.5 U/mL HRP and 75 μ g/mL Amplex Red final concentrations) following activity at 563 nm ($\epsilon_{563} = 56000 \text{ M}^{-1} \text{ cm}^{-1}$).

HPLC analysis and optical rotation: The enantioselectivity of LanDo variant was analyzed by chiral HPLC with equipment consisting of a tertiary pump (Varian/Agilent Technologies) coupled to an autosampler (Merck Millipore) and a Lux 5 μ m Cellulose-1 column (Phenomenex). For the mobile phase hexane and isopropanol in a ratio 9:1 was used. The separation of the enantiomers was performed at a flow rate of 1.0 mL/min. The rotary polarization was measured with a JASCO P-2000 polarimeter. After full conversion, a reaction mixture with the remaining alcohol (1 mg/mL of 1-(*p*-methoxyphenyl)-ethanol) was extracted with ethyl acetate and dissolved in methanol. The measurement was made at 25 °C with a sodium lamp at 589 nm.

Protein modeling: A structural model of the AAO from *P. eryngii* crystal structure at a resolution of 2.55 Å (Protein Data Bank Europe [PDB] accession number 3FIM,^[16] was used as scaffold for the wild type AAO model and the homology models for different mutants were made from 3FIM by PyMol (Schrodinger LLC.; <http://www.pymol.org>).

DNA sequencing: All genes were verified by DNA sequencing (BigDye Terminator v3.1 Cycle Sequencing Kit) using the following primers: primers sense, RMLN and AAOsec1F and primers antisense RMLC, AAOsec2R, and AAOsec3R.

Computational Methods

The diffusion and binding of the secondary alcohol (S)-1-(*p*-methoxyphenyl)-ethanol were studied using the new adaptive-PELE (Protein Energy Landscape Exploration) software.^[17] This adaptive protocol offers improved sampling by running multiple and consecutive short PELE simulations (epochs), setting the initial conditions of each epoch following a reward function that favours sampling of unexplored areas. Even though the ligands were parameterized based on the OPLS2005 force field, the electrostatic charges were derived from the electrostatic potential obtained through quantum calculations at the M06/6-31G* level with Jaguar,^[18] and its rotamer library was built with MacroModel.^[19] Similarly, the FAD cofactor was optimized with a mixed QM/MM calculation at the same level of theory using Qsite.

A single adaptive-PELE simulation was enough to explore the diffusion and binding of (S)-1-(*p*-methoxyphenyl)-ethanol for the native protein and the different variants. The ligand was initially placed between residues Gly52 and Asn56 in all cases. Each simulation used 192 processors, producing 60 epochs of 20 PELE steps each. During the simulation, ligand perturbations (rotations and translations) are coupled with the backbone perturbation to allow protein flexibility. This is achieved by using an anisotropic network model^[20] applied to every Ca atom, while all sidechain conformations within 6 Å of the ligand were predicted each step. An epsilon value of 0.2 was used in the adaptive protocol, meaning that 20% of the processors started each epoch from the structure with the best ligand-FAD distance.

Acknowledgements

This research was supported by the EU project FP7-KBBE-2013-7-613549-INDOX, by the Spanish Government projects BIO2016-79106-R-Lignolition, CTQ2016-74959-R (MINECO/FEDER, EU) and by the Comunidad de Madrid project Y2018/BIO4738-EVOCHIMERA.

References

- <https://www.statista.com/statistics/263102/pharmaceutical-market-worldwide-revenue-since-2001>
- a) J. Caldwell, *Hum. Psychopharmacol.* **2001**, *16*, S67-S71; b) B. S. Sekhon, *J. Mod. Med. Chem.* **2013**, *1*, 10–36; c) N. Chhabra, M. L. Aseri, D. Padmanabhan, *Int J App Basic Med Res.* **2013**, *3*, 16–18.
- N. M. Maier, P. Franco, W. Lindner, *J. Chromatogr. A* **2001**, *12*, 906(1-2):3-33.
- a) K. Faber, *Chemistry.* **2001**, *3*, 7(23):5004-10; b) O. Pàmies, J. E. Bäckvall, *Trends Biotechnol.* **2004**, *22*, 130–5; c) N. J. Turner, *Curr. Opin. Chem. Biol.* **2004**, *8*, 114–9.
- a) G. Fantin, M. Fogagnolo, A. Medici, P. Pedrini, S. Poli, M. Sinigaglia, *Tetrahedron Lett.* **1993**, *34*, 883–884; b) Y. L. Li, J. H. Xu, Y. Xu, *J. Mol. Catal. B* **2010**, *64*, 48–52; c) C. Voss, C. Gruber W. Kroutil, *Angew. Chem. Int. Ed.* **2008**, *47*, 741–745.
- T. Matsuda, R. Yamanaka, K. Nakamura, *Tetrahedron: Asymmetry.* **2009**, *20*, 513–557.
- a) A. Ghanem, H. Y. Aboul-Enein, *Chirality.* **2005**, *17*, 1–15; b) S. Qin, Y. Zhao, B. Wu, B. He, *Appl. Biochem. Biotechnol.* **2016**, *180*, 1456–1466.
- a) F. Escalettes, N. J. Turner, *ChemBioChem* **2008**, *14*, 9, 857–60; b) M. M. Musa, K. I. Ziegelmann-Fjeld, C. Vieile, J. G. Zeikus, R. S. Phillips, *J. Org. Chem.* **2007**, *72*, 30–34.
- a) W. Kroutil, H. Mang, K. Edegger K. Faber, *Adv. Synth. Catal.* **2004**, *346*, 125–142; b) E. W. van Hellemond, L. Vermote, W. Koolen, T. Sonke, E. Zandvoort, D. P. Heuts, D. B. Janssen, M. W. Fraaije, *Adv. Synth. Catal.* **2009**, *351*, 1523–1530; c) W. P. Dijkman, C. Binda, M. W. Fraaije, A. Mattevi, *ACS Catal.* **2015**, *5*, 1833–1839.
- F. J. Ruiz-Dueñas, A. T. Martínez, *Microb. Biotechnol.* **2009**, *2*, 164–77.
- P. Ferreira, A. Hernandez-Ortega, B. Herguendas, A. T. Martínez, M. Medina, *J. Biol. Chem.* **2009**, *284*, 24840–24847.
- J. Viña-Gonzalez, D. Gonzalez-Perez, P. Ferreira, A. T. Martínez, M. Alcalde, *Appl. Environ. Microbiol.* **2015**, *81*, 6451–62; b) J. Viña-Gonzalez, K. Elbl, X. Ponte, F. Valero, M. Alcalde, *Biotechnol. Bioeng.* **2018**, *115*, 1666–1674.
- S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z. G. Zhang, D. J. Opperman, M. T. Reetz, J. P. Acevedo, *ACS Synth. Biol.* **2013**, *2*, 83–92.
- D. Gonzalez-Perez, P. Molina-Espeja, E. Garcia-Ruiz, M. Alcalde, *PLoS One.* **2014**, *9*, e90919.
- F. J. Ruiz Dueñas, P. Ferreira, M. J. Martínez, A. T. Martínez, *Prot. Exp. Purif.* **2006**, *45*, 191–199.
- I. S. Fernandez, F. J. Ruiz-Dueñas, E. Santillana, P. Ferreira, M. J. Martínez, A. T. Martínez, A. Romero, *Acta Crystallogr. Sect. C* **2009**, *65*, 1196–1205.
- D. Lccina, J. F. Gilabert, V. Guallar, *Sci. Rep.* **2017**, *7*, 8466.
- A. D. Bochevarov, E. Harder, T. F. Hughes, J. R. Greenwood, D. A. Braden, D. M. Philipp, D. Rinaldo, M. D. Halls, J. Zhang, R. A. Friesner, *Int. J. Quantum Chem.* **2013**, *113*, 2110–2142.
- a) R. B. Murphy, D. M. Philipp, R. A. Friesner, *J. Comput. Chem.* **2000**, *21*, 1442–1457; b) G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221–234.
- A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, I. Bahar, *Biophys. J.* **2001**, *80*, 505–515.
- A. Hernandez-Ortega, P. Ferreira, A. T. Martínez, *Appl. Microbiol. Biotechnol.* **2012**, *93*, 1395–1410.
- A. Serrano, F. Sancho, a J. Viña-González, J. Carro, M. Alcalde, V. Guallar, A. T. Martínez, *Catal. Sci. Technol.* **2019**, *9*, 833–841.
- A. Hernandez-Ortega, P. Ferreira, P. Merino, M. Medina, V. Guallar, A. T. Acta Crystallogr. Sect. DMartinez, *ChemBioChem.* **2012**, *13*, 427–435.
- a) P. Ferreira, F. J. Ruiz-Dueñas, M. J. Martínez, W. J. H. van Berkel, A. T. Martínez, *FEBS J.* **2006**, *273*, 4878–4888; b) A. Hernandez-Ortega, F. Lucas, P. Ferreira, M. Medina, V. Guallar, A. T. Martínez, *Biochemistry.* **2012**, *51*, 6595–6608; c) J. Carro, P. Amengual-Rigo, F. Sancho, M. Medina, V. Guallar, P. Ferreira, A. T. Martínez, *Sci. Rep.* **2018**, *8*, 8121.
- J. Carro, E. Fernandez-Fueyo, C. Fernandez-Alonso, J. Cañada, R. Ullrich, M. Hofrichter, M. Alcalde, P. Ferreira, A. T. Martínez, *Biotechnol. Biofuels* **2018**, *11*, 86.
- P. Ferreira, M. Medina, F. Guillén, M. J. Martínez, W. J. Van Berkel, A. T. Martínez, *Biochem. J.* **2005**, *1*, 731–8.



Supporting Information

© Copyright Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, 2019

Structure-Guided Evolution of Aryl Alcohol Oxidase from *Pleurotus eryngii* for the Selective Oxidation of Secondary Benzyl Alcohols

Javier Viña-Gonzalez, Diego Jimenez-Lalana, Ferran Sancho, Ana Serrano, Angel T. Martinez, Victor Guallar, and Miguel Alcalde*

Supporting information for:

**Structure-Guided Evolution of Aryl Alcohol Oxidase
from *Pleurotus eryngii* for the Selective Oxidation of
Secondary Benzyl Alcohols**

Javier Viña-Gonzalez¹, Diego Jimenez-Lalana¹, Ferran Sancho², Ana Serrano³, Angel T. Martinez³, Victor Guallar^{2,4} and Miguel Alcalde^{1*}

¹Department of Biocatalysis, Institute of Catalysis, CSIC, Cantoblanco, 28049 Madrid, Spain.

²Barcelona Supercomputing Center, Jordi Girona 31, 08034 Barcelona, Spain.

³Biological Research Center, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain.

⁴ICREA, Passeig Lluís Companys 23, 08010 Barcelona, Spain.

*Corresponding author: malcalde@icp.csic.es

Table of contents

Supporting Table S1

Supporting Figure 1

Supporting Figure 2

Supporting Figure 3

Supporting Table S2

Supporting Table 1. List of primers

<u>Oligo</u>	<u>Sequence</u>
RMLN	5'-CCTCTATACTTTAACGTCAAGG-3'
RMLC	5'-GGGAGGGCGTGAATGTAAGC-3'
I500Af	5'-GAGACAACGCCAACACGGCTTCCACCCAGTTGGAACGGC-3'
I500Ar	5'-GCCGTTCCAACCTGGGTGGAAAGCCGTGTTGGCGTTGTCTC-3'
F501Af	5'-GAGACAACGCCAACACGATTGCTCACCCAGTTGGAACGGC-3'
F501Ar	5'-GCCGTTCCAACCTGGGTGAGCAA7CGTGTGGCGTTGTCTC-3'
DM5001Af	5'-GAGACAACGCCAACACGGCTGCTCACCCAGTTGGAACGGC-3'
DM5001Ar	5'-GCCGTTCCAACCTGGGTGAGCAGCCGTGTTGGCGTTGTCTC-3'
22c1F	5'-GAGACAACGCCAACACGNDTNDTCACCCAGTTGGAAC-3'
22c1R	5'-GTTCCAACCTGGGTGAHNAHNCGTGTTGGCGTTGTCTC-3'
22c2F	5'-GAGACAACGCCAACACGNDTVHGCACCCAGTTGGAAC-3'
22c2R	5'-GTTCCAACCTGGGTGCDBAHNCGTGTTGGCGTTGTCTC-3'
22c3F	5'-GAGACAACGCCAACACGNDTTGGCACCCAGTTGGAAC-3'
22c3R	5'-GTTCCAACCTGGGTGCCAAHNCGTGTTGGCGTTGTCTC-3'
22c4F	5'-GAGACAACGCCAACACGVHGNITCACCCAGTTGGAAC-3'
22c4R	5'-GTTCCAACCTGGGTGAHNCDBCCTGTTGGCGTTGTCTC-3'
22c5F	5'-GAGACAACGCCAACACGVHGVHGCACCCAGTTGGAAC-3'
22c5R	5'-GTTCCAACCTGGGTGCDDBCCTGTTGGCGTTGTCTC-3'
22c6F	5'-GAGACAACGCCAACACGVHGTGGCACCCAGTTGGAAC-3'
22c6R	5'-GTTCCAACCTGGGTGCCACDBCCTGTTGGCGTTGTCTC-3'
22c7F	5'-GAGACAACGCCAACACGTGGNDTCACCCAGTTGGAAC-3'
22c7R	5'-GTTCCAACCTGGGTGAHNCCACGTGTTGGCGTTGTCTC-3'
22c8F	5'-GAGACAACGCCAACACGTGGVHGCACCCAGTTGGAAC-3'
22c8R	5'-GTTCCAACCTGGGTGCDCCACCTGTTGGCGTTGTCTC-3'

III. Results

22c9F	5'-GAGACAACGCCAACACG <i>TGGTGGC</i> ACCCAGTTGGAAC-3'
22c9R	5'-GTTCCAAC <i>TGGGTGCC</i> ACCACGTGTTGGCGTTGTCTC-3'
RHF1	5'-GGAACAAGGCCGGGCGCAAG-3'
FM1	5'-TAGGGGCAGAGGCTCCACTC-3'
RM1	5'-GCATAGCGATCGAAATCTTC-3'
FHF-1	5'-TCATGATGCGTGGATCAACA-3'
RHF-2	5'-AGGAGCAAATGGTCGGATAG-3'
FM2	5'-ATCCTAGCGTAGGCCGAAAC-3'
RM2	5'-TCTCCGCGAGCTACAGGAGA-3'
FHF-2	5'-TTATGAGTGTTACAAACGCGTTGATT-3'
RHF-3	5'-GTTGGCGTTGTCTCTAATGTACGACTC-3'
FM3	5'-CGACGGACGATGCTGCTATC-3'
F116	5'-ACGTGTCCGTCTTGGTCCTAGAGGCCGGGTGTATCAGATGRGAA-3
R116	5'-TCTGCCCTAATACATT <i>CYCA</i> -3
F5	5'- <i>CARTTGYGTATCCTYGC</i> GGCCGTA <i>RCTAG</i> GGGGGTCTAGCTCTGYTAA-3
R5	5'-TTARCAGAGCTAGACCCCCCTAGYATACGGCCGCRAGGATACRCAAYTG-3
F521	5'- GCGTCATGGCCACGACGCRAGAGCAAA-3
R521	5'- GTCGGGATTGAAGAAGAACTCTTCGCTTTGCTCTYGCCTCGTGG -3
F9419	5'-ATAACATCYTCAGAGACTCGTCCGAGTTCAACGYTGATTTA-3
R9419	5'-TAAATCARCGTTGAACTCGGACGAGTCTCTGARGATGTTAT-3
F1397	5'-TTCGTTTCCTCTCTGGT <i>CRAGC</i> -3
R1397	5'-GTATAACGAAGTCCGCCACGCTY <i>GACC</i> -3
AAOsec1F	5'-GTGGATCAACAGAAGATTTTCGATCG-3'
AAOsec2R	5'-GTGGTTAGCAATGAGCGCGG-3'
AAOsec3R	5'-GGAGTCGAGCCTCTGCCCT-3'

Codon substitutions are shown in italics (where N = A/T/C/G; D = no C; V =no T, H = no G; B = no A; R = A/G; Y = C,T).

Supporting Table 2. Polarimeter measurements for the LanDo variant reaction with 1-(*p*-methoxyphenyl)-ethanol (correspondent to the remaining R-enantiomer). Number of cycles: 8 with 1 sec cycle interval. Path length 100mm.

PMT Voltage[V]	Temperature[°C]	Optical Rotation Monitor
353	24.96	0.0592
353	24.95	0.0601
353	24.96	0.0593
354	24.96	0.0602
353	24.98	0.0586
353	24.99	0.0603
354	25.01	0.0600
354	25.03	0.0589

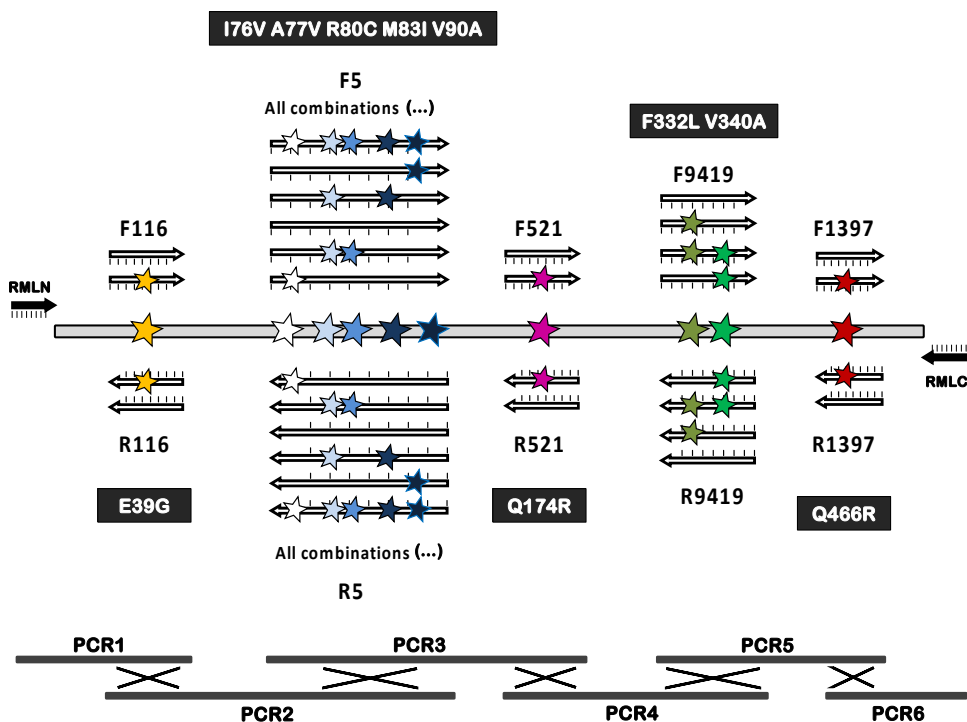


Figure S1. Method for *in vivo* site-directed recombination. Primers designed to be used in the site-directed recombination experiment for the PCR amplification of the selected mutated positions (in black). For each mutation, adjacent sense and antisense primers were synthesized that were 50% mutated at the sites of interest. Six PCR reactions were performed with ~40 bp homologous sequences at each end to foster *in vivo* recombination (7 crossover events). The PCR fragments were assembled by transformation into yeast with the linearized vector to yield a library of all combinations of the mutations/reversions in one-pot .

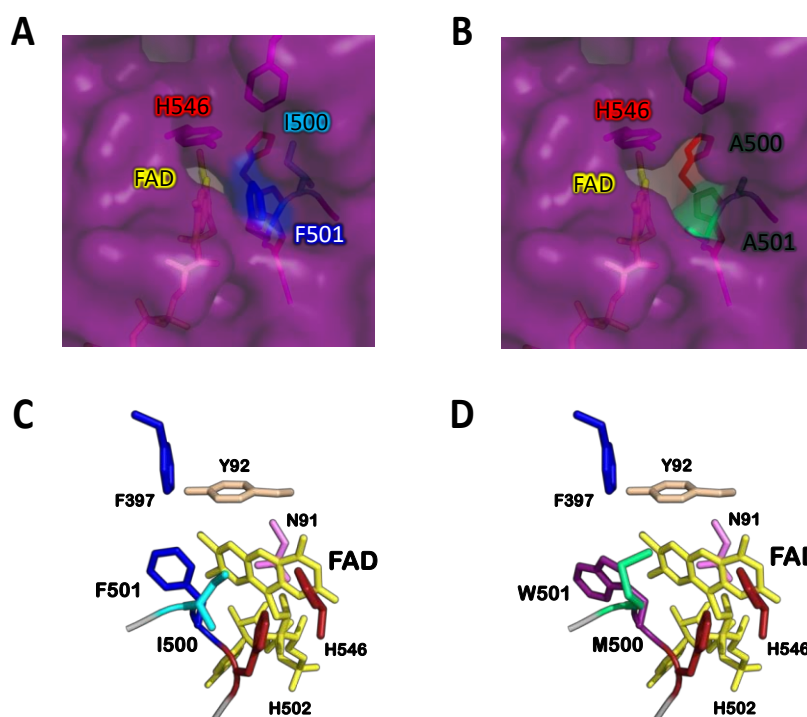


Figure S2. The access channel and catalytic pocket before and after mutation. The channel giving access to the active site in the FX9 (**A**) and I500A-F501A (**B**) mutants: the FAD molecule is depicted in yellow, the Phe501 and Ile500 residues in the parental type are in blue and light blue, the Ala500 and Ala501 mutations are in green, and histidine 546 at the catalytic pocket is in red. Catalytic pocket in the FX9 (**C**) and 15G12 (**D**) variants: His502 and His546 at the active site are depicted in red, FAD is depicted in yellow, ancestral/consensus mutation Asn91 is in pink, Phe397 and Phe501 are in blue, Ile500 in light blue, Tyr92 is depicted in white, whereas the new Trp501 and Met500 substitutions are depicted in purple and light green, respectively. The models were prepared from the crystal structure of *P. eryngii* AAO (PDB 3FIM).

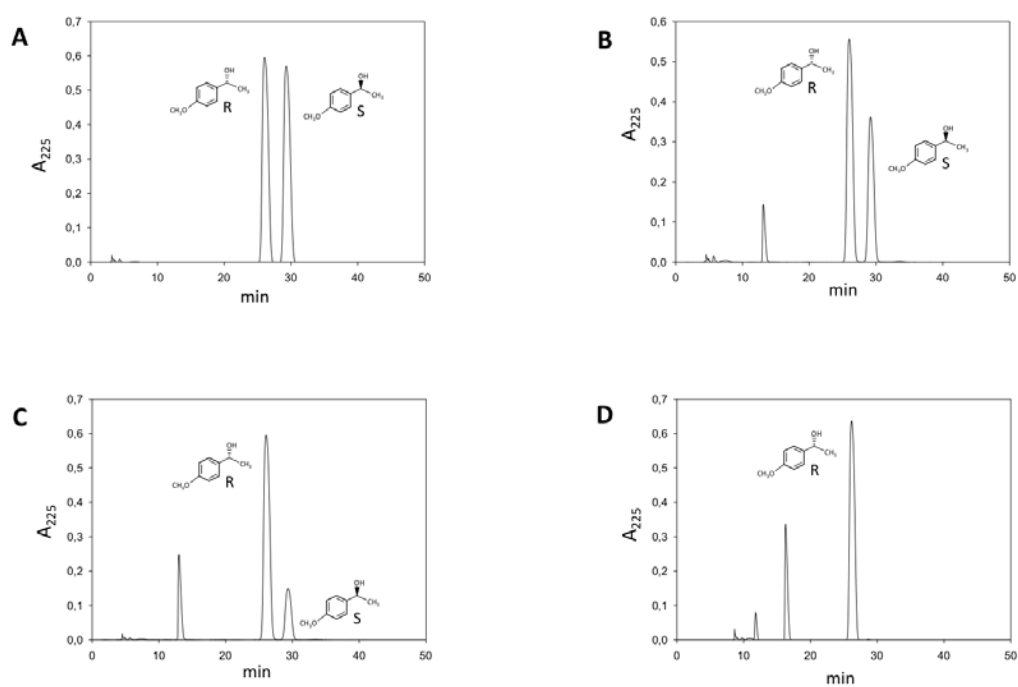


Figure S3. Chiral HPLC chromatograms. Elution profiles of the reaction of LanDo variant (1 μM) with racemic 1-(*p*-methoxyphenyl)-ethanol (2.5 mM). Negative control (A), after 15 minutes reaction (B), after 45 min reaction (C) and after 90 min reaction (D).

1.4. Article IV

Towards the Efficient Production of FDCA: Insights from Molecular Modeling of AaeUPO, HMFO, and AAO enzymes

Marina Cañellas,^{‡,1} Ferran Sancho,^{‡,2} Victor Guallar^{2,3} and Fátima Lucas^{4,*}

¹ Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, ON K1N 6N5, Canada

² Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain

³ Icrea, Passeig Lluís Companys 23, E-08010 Barcelona, Spain

⁴ ZYMVOL BIOMODELING SL, C/ Roc Boronat, 117, 08018 Barcelona, Spain

* Correspondence: flucas@zymvol.com

‡ These two authors equally contributed to the work

Towards the Efficient Production of FDCA: Insights from Molecular Modeling of *Aae*UPO, HMFO, and AAO enzymes.

Marina Cañellas,^{‡,1} Ferran Sancho,^{‡,2} Victor Guallar^{2,3} and Fátima Lucas^{4,*}

¹ Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, ON K1N 6N5, Canada

² Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain

³ ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain

⁴ ZYMVOL BIOMODELING SL, C/ Roc Boronat, 117, 08018 Barcelona, Spain

* Correspondence: flucas@zymvol.com

‡ These authors equally contributed to this work

ABSTRACT The growing interest in replacing fossil-based products with renewable chemical building blocks has focused researchers' 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 cascade reaction 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 a potentially improved *Aae*UPO towards the enzymatic production of FDCA.

SIGNIFICANCE There is an imperious interest in finding robust biocatalysts to produce FDCA from HMF. The present study makes valuable contributions to the field by giving a complete biophysical description of a number of enzymatic cascades that lead to the production of FDCA in three different enzymes (*Aae*UPO, AAO, and HMFO). By the use of molecular modeling techniques, we have been able to dissect the reasons underneath the low or non-activity at particular steps of the reaction that were hindering FDCA production. All the gathered information not only has enabled us to computationally engineer the *Aae*UPO enzyme towards improved FDCA production but also opens the door to further attempts of tailoring these three now well-characterized enzymes.

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 part of the large amount of gasoline and diesel used in transport worldwide (~14,2 millions of barrels/day in 2018) (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 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 2,5-furandicarboxylate (PEF) polymer (9-11), 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 that render the process polluting and expensive (12, 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 (15). 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-hydroxymethylfuran carboxylic acid (HMFC). Both metabolites are precursors of 2,5-formylfuran carboxylic 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 (13, 16, 17). 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, named HMF oxidase (HMFO), has been described to

oxidize HMF giving FDCA as the only product (13). However, HMFO enzyme shows poor activity towards FFCA intermediate, limiting the overall process' efficiency (18, 19). 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 (20).

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* (21). 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 HMFO, including the FFCA oxidation, although 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 H₂O₂, which is then used by the UPO to produce FDCA from FFCA. In this protocol, H₂O is the only by-product formed and O₂ 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 behind 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 HMFO 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 (22), 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 enabled us to perform the *in silico* engineering of UPO enzyme towards the theoretical improvement of FDCA production, by using both rational design and an *in silico* evolution protocol (23).

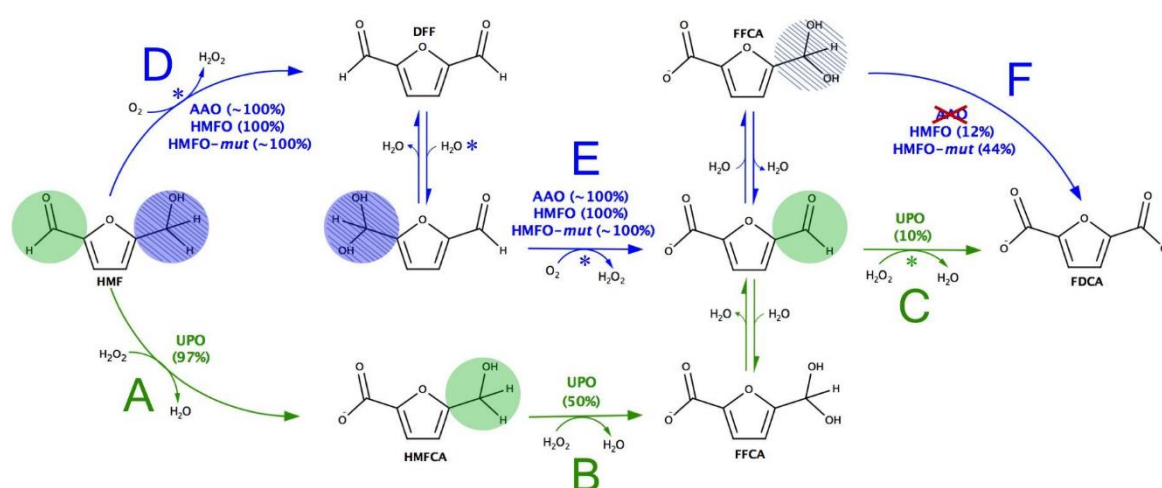


Figure 1. Schematic representation of HMF conversion into FDCA by two different routes: via DFF (catalyzed by AAO, HMFO and HMFO-mut enzymes, colored in blue) and via HMFCFA (catalyzed by UPO enzyme, colored in green). Functional groups that are oxidized by UPO in each step of the reaction are circled in green, while functional groups oxidized by AAO are circled in blue and blue-striped for HMFO/HMFO-mut. Reaction steps marked with an asterisk are those involved in the protocol proposed by Carro et al. to obtain FDCA (21). For each step, the conversion rate is shown considering HMF as the reactant in a 24h 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 (24). Crystal structures with PDB code 3FIM and 4UDP were used respectively for simulations with AAO and HMFO. Simulations were performed at pH 7, reproducing the experimental conditions, which has been found to be a good compromise for the three proteins in terms of performance and stability (21). Protein structures were prepared accordingly using Schrodinger's Protein Preparation Wizard PROPKA module (25-27) and H++ web server (<http://biophysics.cs.vt.edu/>) (28). 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 fully 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 (30) with the density functional M06 (31) with 6-31G** basis set and a Poisson Boltzmann Finite (PBF) implicit solvent (32) in order to

obtain their electrostatic potential derived atomic charges. These include HMF, DFF, FFCA (deprotonated, $pK_a=3.09$) and HMFCA (deprotonated, $pK_a= 3.11$). Note that for AAO and HMFO simulations, FFCA and DFF were modeled on their hydrated form. All pK_a values were collected from www.chemicalize.org (33). 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 (34). 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

To investigate the differences observed experimentally on yield and regioselectivity between the distinct reactions, we have used Protein Energy Landscape Exploration (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 (35) and a surface generalized Born implicit solvent (36). 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 (37-39).

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 was 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. (23) and successfully applied for the design of POXA1b laccase by Giacobelli et al. (40). 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 mutated to all the possible remaining 19 amino acids. Mutated residues included positions: 69, 77, 76, 121, 123, 126, 188, 189, 190, 191, 192, 193, 196 and 199. After mutation, the ligand was perturbed with 0.1 Å translations and 0.01 rad rotations whereas the protein backbone was subjected to 0.5 Å ANM displacements, and all the residues within 5 Å from the substrate were energy minimized. For each mutant, a 4-h simulation was performed, obtaining as a result the enzyme-substrate interaction energy, that was used to rank the 266 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 80 48h trajectories.

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 efficient production of FDCA from FFCA in an attempt to theoretically enhance the whole reaction.

Unspecific peroxygenase via HMFCA – reactions A, B, and C.

Reaction A. According to the results obtained by Carro et al. (21), 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 is 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 6 Å 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 H₂ and H₃, 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 (H₂, H₃, and H₆) and the O-Heme in structures corresponding to this minimum. Results show that the average distance between H₆ and O-Heme is 1.77 Å, compared to 4.27 Å for H₂ and 3.55 Å for H₃, providing evidence that support our hypothesis. Therefore, we propose that, because 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 inspection 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 positioning leads to a reduction of the diameter of the heme channel entrance (supplementary 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 inhibition. This is in a good agreement with experimental results, which show a decreased oxidation rate of HMF over time (72% in 8h versus 97% in 24h) (21).

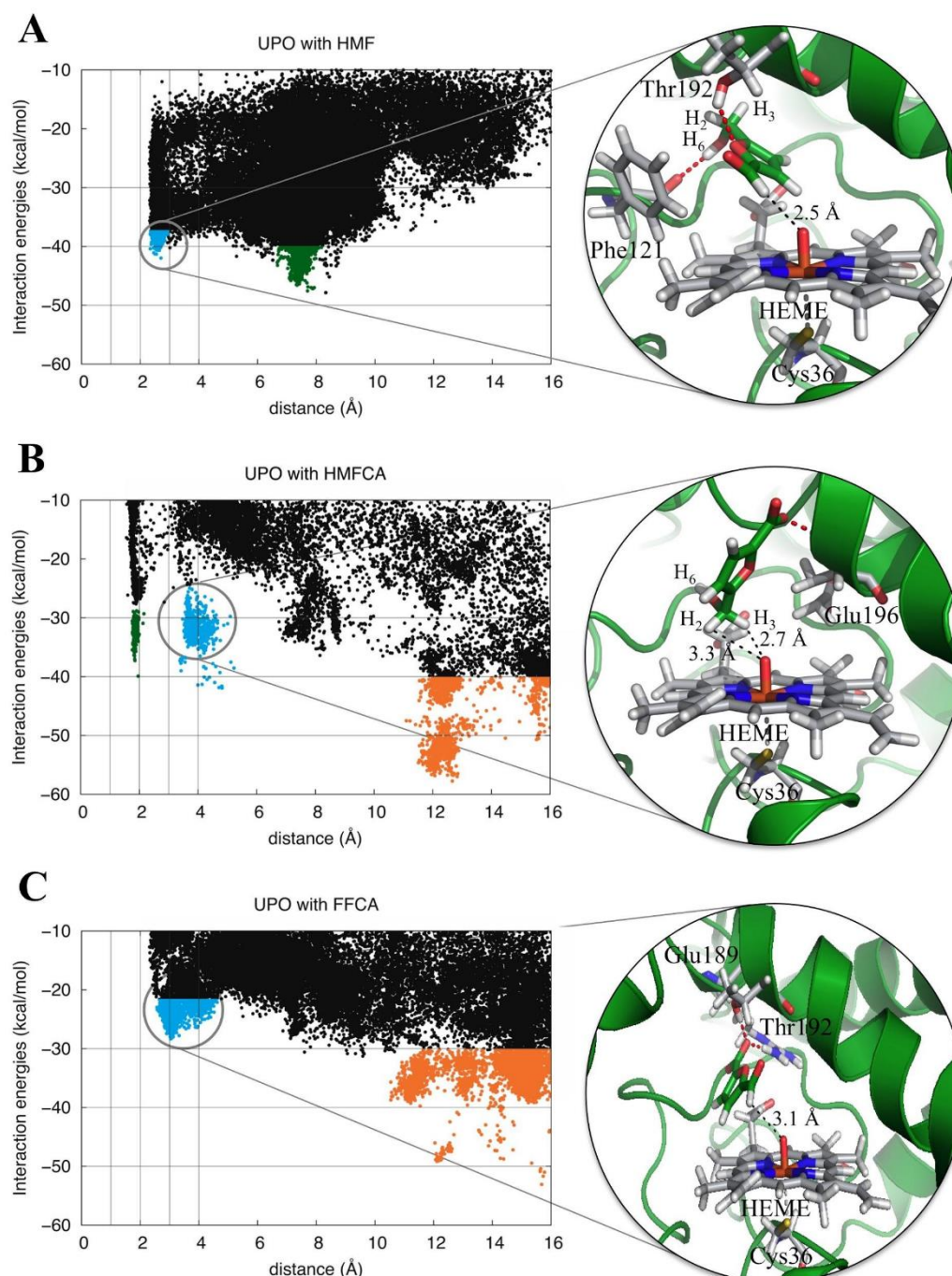


Figure 2. The plots on the left are the representation of the interaction energy of each PELE structure against its reactive distance in UPO from *A. aegerita* for HMF (A), HMFCFA (B) and FFCA (C). This distance (Å) is measured between the reactive O atom in the H₂O₂-activated heme (compound I) and reactive 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.

Despite the narrowed heme path entrance, the substrate is still able to enter the heme site and adopt favorable orientations for the reaction. As shown in **Fig. 2B**, there are two different minima around -35 kcal/mol, both 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 H₂ and H₃ 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 H₂ and H₃ atoms are able to approach O-Heme atom within an appropriate distance to react. See **Fig. 2B** for a molecular representation of this minimum.

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 important 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 two highlighted mutations: Phe121Lys and Ala77Ser. 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 Phe121Lys mutation in the binding pocket. As shown in **Fig. 3**, the new variant 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 (<2.5 Å), along with a smoother entrance to the binding site.

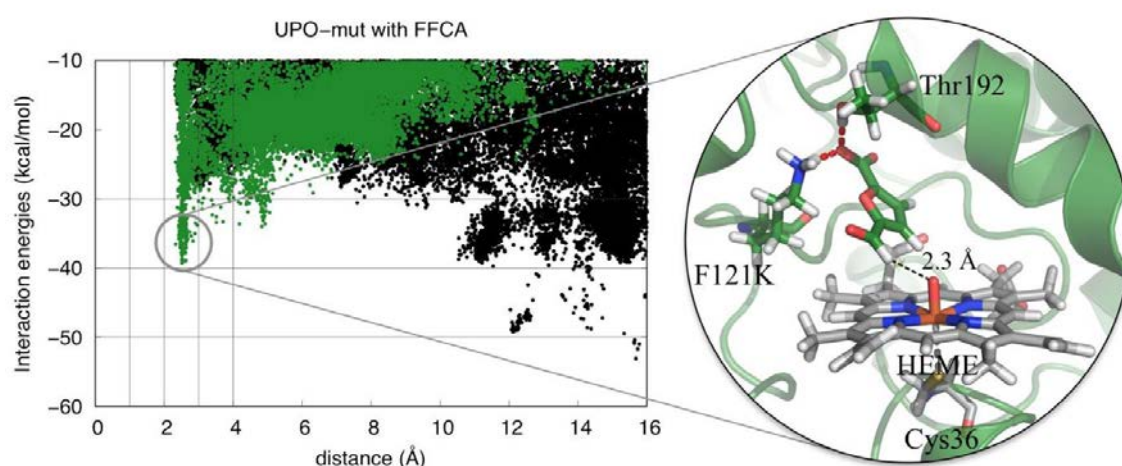


Figure 3. UPO variant with FFCA substrate. On the left, there is a representation of the interaction energy of each PELE structure against its reactive distance for the wt UPO (colored in black) and for the UPO variant (colored in green). Reactive distance (Å) is measured between the H₂O₂-activated heme (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 are 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.

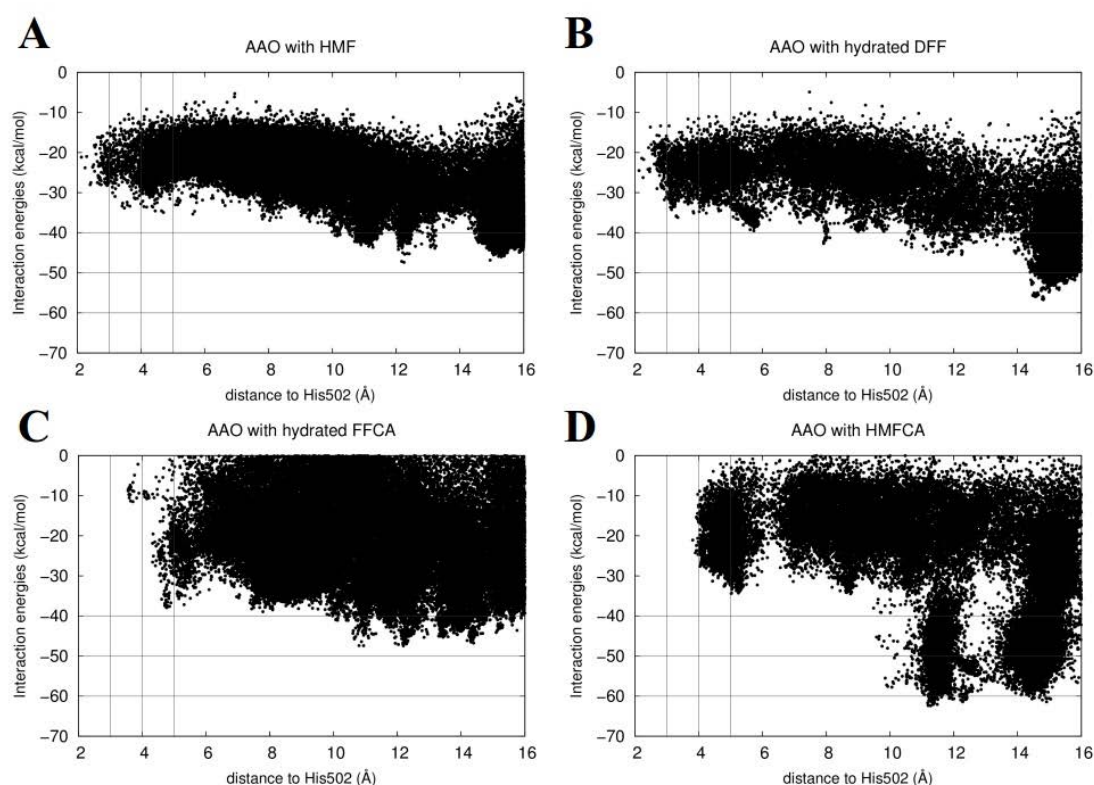


Figure 4. Interaction energies vs. ligand distances from PELE simulations in AAO from *Pleurotus eryngii*. The distances shown (Å) are measured between the reactive N atom in the His502 and the alcohol hydrogen atom from the ligand. The plots correspond to the migration of different ligands: **A** - HMF, **B** - DFF, **C** - FFCA, and **D** - HMFA.

Aryl alcohol oxidase via DFF – reactions D, E, and F.

As explained in the introduction, it 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 ($\sim 100\%$ 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, opening the possibility for tailoring the protein to overcome this limitation.

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. Thus, the distance between the hydrogen from the ligand's alcohol and His502 nitrogen is the reaction coordinate to consider. The interaction energy along the simulations has been

represented in **Fig. 4** as a function of this distance. A threshold of 3 Å was established in order to consider that the ligand is close enough for the proton abstraction to take place. As shown in **Fig. 3**, both HMF and DFF reach distance values lower than 3 Å, which are appropriate for the reaction, while FFCA cannot get closer than 4 Å. Based on this difference, we hypothesize that FFCA is not converted by AAO since 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 which has been proven experimentally to be non-reactive in AAO, were also performed. Interestingly, its energy profile shows similar results to FFCA, stalling at 4 Å 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 4 Å for both ligands have been extracted, and a representative conformation for FFCA ligand is shown in **Fig. 5**. 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 responsible for the null activity of both ligands in AAO.

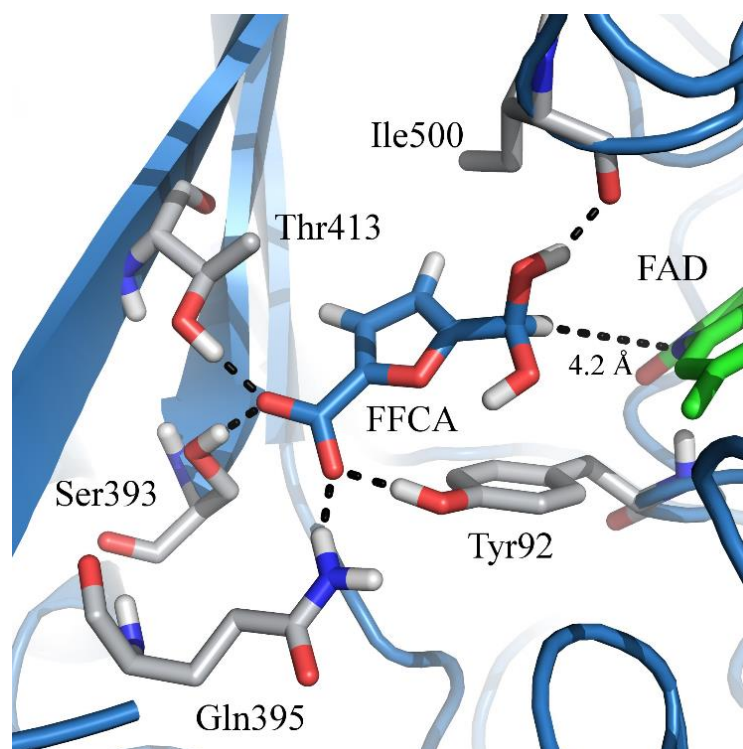


Figure 5. Structural representation of the FFCA ligand conformation when the catalytic distance is around 4 Å. Strong interactions of the carboxylic group of the ligand with polar residues can be appreciated.

Hydroxymethylfurfural oxidase 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.

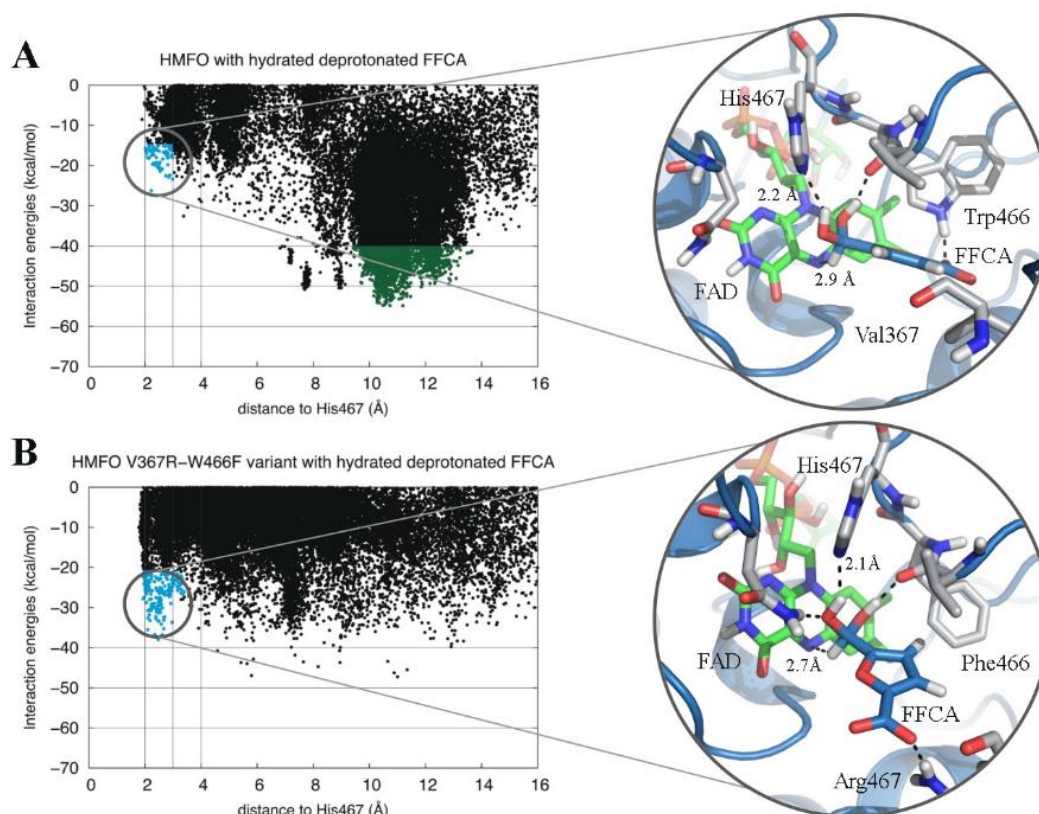


Figure 6. Two-dimensional interaction maps showing the interaction energy vs. the distance between FFCA and the His467 for the native HMFO (A) and HMFO-*mut* (B). A representative structure from the catalytic minimum was taken from the native protein where it can be seen the interaction between Trp466 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. 6. It is shown how for the native protein (Fig. 6A), the ligand is able to contact the active site with reasonable catalytic distances for the oxidation to occur ($< 3 \text{ \AA}$), but the catalytic position is not energetically favored. Around 10 \AA 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. 6B**). 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. 6**, 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 in finding bio-based alternatives for the production of FDCA from HMF. 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 a few systems have been proven to perform the whole transformation, and in most of the cases, the reaction from FFCA to FDCA has still a really poor conversion. Here, the study of these enzymatic transformations using molecular modeling 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 modeling simulations and to the employment of a new methodology for *in silico* enzyme evolution, we have been able to propose a computationally engineered 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. 2012. Conversion of biomass to selected chemical products. *Chem Soc Rev* 41(4):1538-1558.
2. Huber, G. W., S. Iborra, and A. Corma. 2006. Synthesis of transportation fuels from biomass: chemistry, catalysts, and engineering. *Chem Rev* 106(9):4044-4098.
3. Corma, A., S. Iborra, and A. Velty. 2007. Chemical routes for the transformation of biomass into chemicals. *Chem Rev* 107(6):2411-2502.
4. 2018. International Petroleum and Other Liquids Production, Consumption, and Inventories. U.S. Department of Energy, Energy Information Administration, Independent Statistics & Analysis.
5. Bozell, J. J., and G. R. Petersen. 2010. Technology development for the production of biobased products from biorefinery carbohydrates-the US Department of Energy's "Top 10" revisited. *Green Chemistry* 12(4):539-554. 10.1039/B922014C.
6. Wierckx, N., T. Elink Schuurman, L. Blank, and H. Ruijsenaars. 2015. Whole-Cell Biocatalytic Production of 2,5-Furandicarboxylic Acid. In *Microbiology Monographs. Microorganisms in Biorefineries*. B. Kamm, editor. Springer Berlin Heidelberg, pp. 207-223.
7. Gallezot, P. 2007. Process options for converting renewable feedstocks to bioproducts. *Green Chemistry* 9(4):295-302. 10.1039/B615413A.
8. Werpy, T., and G. Petersen. 2004. 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*. Medium: ED; Size: 76 pp. pages.
9. Jong, E. d., M. A. Dam, L. Sipos, and G. J. M. Gruter. 2012. Furandicarboxylic Acid (FDCA), A Versatile Building Block for a Very Interesting Class of Polyesters. In *ACS Symposium Series, vol. 1105. Biobased Monomers, Polymers, and Materials*. American Chemical Society, pp. 1-13.
10. Koopman, F., N. Wierckx, J. H. de Winde, and H. J. Ruijsenaars. 2010. Efficient whole-cell biotransformation of 5-(hydroxymethyl)furfural into FDCA, 2,5-furandicarboxylic acid. *Bioresour Technol* 101(16):6291-6296.
11. Papageorgiou, G. Z., V. Tsanaktis, and D. N. Bikiaris. 2014. Synthesis of poly(ethylene furandicarboxylate) polyester using monomers derived from renewable resources: thermal behavior comparison with PET and PEN. *Phys Chem Chem Phys* 16(17):7946-7958.

12. van Putten, R. J., J. C. van der Waal, E. de Jong, C. B. Rasrendra, H. J. Heeres, and J. G. de Vries. 2013. Hydroxymethylfurfural, a versatile platform chemical made from renewable resources. *Chem Rev* 113(3):1499-1597.
13. Sajid, M., X. Zhao, and D. Liu. 2018. Production of 2,5-furandicarboxylic acid (FDCA) from 5-hydroxymethylfurfural (HMF): recent progress focusing on the chemical-catalytic routes. *Green Chemistry* 20(24):5427-5453. 10.1039/C8GC02680G.
14. Thomas, S. M., R. DiCosimo, and V. Nagarajan. 2002. Biocatalysis: applications and potentials for the chemical industry. *Trends Biotechnol* 20(6):238-242.
15. Carlini, C., P. Patrono, A. M. R. Galletti, G. Sbrana, and V. Zima. 2005. Selective oxidation of 5-hydroxymethyl-2-furaldehyde to furan-2,5-dicarboxaldehyde by catalytic systems based on vanadyl phosphate. *Applied Catalysis A, General* 289(2):197-204.
16. Hanke, P. D. 2009. Enzymatic oxidation of HMF. Google Patents.
17. Koopman, F., N. Wierckx, J. H. de Winde, and H. J. Ruijsenaars. 2010. Identification and characterization of the furfural and 5-(hydroxymethyl)furfural degradation pathways of *Cupriavidus basilensis* HMF14. *Proc Natl Acad Sci U S A* 107(11):4919-4924.
18. Dijkman, W. P., and M. W. Fraaije. 2014. Discovery and characterization of a 5-hydroxymethylfurfural oxidase from *Methylovorus* sp. strain MP688. *Appl Environ Microbiol* 80(3):1082-1090.
19. Dijkman, W. P., D. E. Groothuis, and M. W. Fraaije. 2014. Enzyme-catalyzed oxidation of 5-hydroxymethylfurfural to furan-2,5-dicarboxylic acid. *Angew Chem Int Ed Engl* 53(25):6515-6518.
20. Dijkman, W. P., C. Binda, M. W. Fraaije, and A. Mattevi. 2015. Structure-Based Enzyme Tailoring of 5-Hydroxymethylfurfural Oxidase. *ACS Catalysis* 5(3):1833-1839.
21. Carro, J., P. Ferreira, L. Rodríguez, A. Prieto, A. Serrano, B. Balcells, A. Ardá, J. Jiménez-Barbero, A. Gutiérrez, R. Ullrich, M. Hofrichter, and A. T. Martínez. 2015. 5-hydroxymethylfurfural conversion by fungal aryl-alcohol oxidase and unspecific peroxygenase. *FEBS J* 282(16):3218-3229.
22. Borrelli, K. W., A. Vitalis, R. Alcantara, and V. Guallar. 2005. PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *Journal of Chemical Theory and Computation* 1(6):1304-1311.
23. Monza, E., M. F. Lucas, S. Camarero, L. C. Alejaldre, A. T. Martínez, and V. Guallar. 2015. Insights into Laccase Engineering from Molecular Simulations: Toward a Binding-Focused Strategy. *The Journal of Physical Chemistry Letters* 6(8):1447-1453.

24. Piontek, K., E. Strittmatter, R. Ullrich, G. Gröbe, M. J. Pecyna, M. Kluge, K. Scheibner, M. Hofrichter, and D. A. Plattner. 2013. Structural basis of substrate conversion in a new aromatic peroxygenase: cytochrome P450 functionality with benefits. *J Biol Chem* 288(48):34767-34776.
25. Madhavi Sastry, G., M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design* 27(3):221-234.
26. Søndergaard, C. R., M. H. M. Olsson, M. Rostkowski, and J. H. Jensen. 2011. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *Journal of Chemical Theory and Computation* 7(7):2284-2295.
27. Olsson, M. H. M., C. R. Søndergaard, M. Rostkowski, and J. H. Jensen. 2011. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of Chemical Theory and Computation* 7(2):525-537.
28. Anandkrishnan, R., B. Aguilar, and A. V. Onufriev. 2012. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research* 40(Web Server issue):W537-W541.
29. Hernández-Ortega, A., K. Borrelli, P. Ferreira, M. Medina, Angel T. Martínez, and V. Guallar. 2011. Substrate diffusion and oxidation in GMC oxidoreductases: an experimental and computational study on fungal aryl-alcohol oxidase. *Biochemical Journal* 436(2):341.
30. Jaguar, version 8.4, Schrödinger, LLC, New York, NY, 2014.
31. Zhao, Y., and D. G. Truhlar. 2006. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *The Journal of Chemical Physics* 125(19):194101.
32. Marten, B., K. Kim, C. Cortis, R. A. Friesner, R. B. Murphy, M. N. Ringnalda, D. Sitkoff, and B. Honig. 1996. New Model for Calculation of Solvation Free Energies: Correction of Self-Consistent Reaction Field Continuum Dielectric Theory for Short-Range Hydrogen-Bonding Effects. *The Journal of Physical Chemistry* 100(28):11775-11788.
33. Swain, M. 2012. chemicalize.org. *Journal of Chemical Information and Modeling* 52(2):613-615.
34. Qsite, version 6.3, Schrödinger, LLC, New York, NY, 2014.
35. Kaminski, G. A., R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. 2001. 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* 105(28):6474-6487.
36. Bashford, D., and D. A. Case. 2000. Generalized born models of macromolecular solvation effects. *Annu Rev Phys Chem* 51:129-152.
 37. Lucas, M. F., E. Monza, L. J. Jørgensen, H. A. Ernst, K. Piontek, M. J. Bjerrum, Á. T. Martínez, S. Camarero, and V. Guallar. 2017. Simulating Substrate Recognition and Oxidation in Laccases: From Description to Design. *Journal of Chemical Theory and Computation* 13(3):1462-1467.
 38. Acebes, S., E. Fernandez-Fueyo, E. Monza, M. F. Lucas, D. Almendral, F. J. Ruiz-Dueñas, H. Lund, A. T. Martínez, and V. Guallar. 2016. Rational Enzyme Engineering Through Biophysical and Biochemical Modeling. *ACS Catalysis* 6(3):1624-1629.
 39. Babot, E. D., 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. *Applied and environmental microbiology* 81(12):4130-4142.
 40. Giacobelli, V. G., E. Monza, M. Fatima Lucas, C. Pezzella, A. Piscitelli, V. Guallar, and G. Sannia. 2017. Repurposing designed mutants: a valuable strategy for computer-aided laccase engineering – the case of POXA1b. *Catalysis Science & Technology* 7(2):515-523. 10.1039/C6CY02410F.

2. Laccases

2.1. Article V

Probing the Surface of a Laccase for Clues towards the Design of Chemo-Enzymatic Catalysts

Viviane Robert,^[a] Emanuele Monza,^[b] Lionel Tarrago,^[a] Ferran Sancho,^[b] Anna De Falco,^[a] Ludovic Schneider,^[a] Eloïne Npetgat Ngoutane,^[a] Yasmine Mekmouche,^[a] Pierre Rousselot Pailley,^[a] A. Jalila Simaan,^[a] Victor Guallar,^[b, c] and Thierry Tron^{*[a]}

[a] Dr. V. Robert, Dr. L. Tarrago, A. De Falco, Dr. L. Schneider, Dr. E. Npetgat Ngoutane, Dr. Y. Mekmouche, Dr. P. R. Pailley, Dr. A. J. Simaan, Dr. T. Tron

Aix Marseille, CNRS, iSm2 UMR 7313

13397 Marseille (France)

E-mail: thierry.tron@univ-amu.fr

[b] E. Monza, F. Sancho, Prof. V. Guallar

Joint BSC-CRG-IRB Research Program in Computational Biology

Barcelona Supercomputing Centre

Jordi Girona 29, 08034 Barcelona (Spain)

[c] Prof. V. Guallar

ICREA

Passeig Lluís Companys 23, 08010 Barcelona (Spain)



Probing the Surface of a Laccase for Clues towards the Design of Chemo-Enzymatic Catalysts

Viviane Robert,^[a] Emanuele Monza,^[b] Lionel Tarrago,^[a] Ferran Sancho,^[b] Anna De Falco,^[a] Ludovic Schneider,^[a] Eloïne Npetgat Ngoutane,^[a] Yasmina Mekmouche,^[a] Pierre Rousselot Pailley,^[a] A. Jalila Simaan,^[a] Victor Guallar,^[b, c] and Thierry Tron^{*,[a]}

In memory of T. Silviu Balaban

Systems featuring a multi-copper oxidase associated with transition-metal complexes can be used to perform oxidation reactions in mild conditions. Here, a strategy is presented for achieving a controlled orientation of a ruthenium–polypyridyl graft at the surface of a fungal laccase. Laccase variants are engineered with unique surface-accessible lysine residues. Distinct ruthenium–polypyridyl-modified laccases are obtained by the reductive alkylation of lysine residues precisely located relative to the T1 copper centre of the enzyme. In none of these

hybrids does the presence of the graft compromise the catalytic efficiency of the enzyme on the substrate 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid). Furthermore, the efficiency of the hybrids in olefin oxidation coupled to the light-driven reduction of O₂ is highly dependent on the location of the graft at the enzyme surface. Simulated Ru^{II}–Cu^{II} electron coupling values and distances fit well the observed reactivity and could be used to guide future hybrid designs.

Introduction

The development of new bio-inspired processes for the sustainable production of chemicals is at the heart of today's grand challenges. The minimal requirements for a (bio)catalyst are: high activity, robustness and ease of production (i.e., low cost). Beyond this scope, specific physicochemical characteristics such as optimal selectivity and activity, solvent tolerance or stability at a given pH are unlikely to be all inherent properties

of an enzyme and are still challenging to obtain in a synthetic catalyst by design. In the case of enzymes, a process-driven customisation is achievable through the combination of molecular evolution techniques and highly efficient expression systems.^[1–3] Furthermore, like the cellular post-translational modifications that increase the functional diversity of the proteome (e.g., the covalent addition of functional groups, the formation of intra- or intermolecular bonds or a proteolytic cleavage), the functionality of recombinant enzymes can be refined using well-established techniques of chemical modification of the enzyme surface.^[4–7] The literature is abundant with examples of proteins covalently modified with tags of all kinds for different purposes: purification, stabilisation, labelling, or even editing enzyme function.^[8–11] Therefore, the combination of surface-located modulators with robust enzymes offers opportunities for developing biomolecule-based catalysts with entirely new sets of functions.

The enzyme laccase (EC1.10.3.2, from plant, fungi, or bacteria) is a robust catalyst largely studied for its potential uses in industrial bleaching processes, organic synthesis applications and in biofuel cells as a cathodic catalyst that reduces dioxygen into water.^[12,13] Laccases couple the oxidation of a wide range of organic and inorganic substrates occurring at a surface-located mononuclear centre (one type 1 Cu^{II} ion) to the four-electron reduction of dioxygen into water, which occurs at a buried trinuclear centre (TNC, structured between a type 3 pair and type 2 Cu^{II} ions) through successive intramolecular electron transfers (ETs).^[14,15] Expanding laccase reactivity through their combination with small and diffusible synthetic redox mediators is well documented.^[16] Recently, we have in-

[a] Dr. V. Robert, Dr. L. Tarrago, A. De Falco, Dr. L. Schneider, Dr. E. Npetgat Ngoutane, Dr. Y. Mekmouche, Dr. P. R. Pailley, Dr. A. J. Simaan, Dr. T. Tron
Aix Marseille Université
Centrale Marseille, CNRS, iSm2 UMR 7313
13397 Marseille (France)
E-mail: thierry.tron@univ-amu.fr

[b] E. Monza, F. Sancho, Prof. V. Guallar
Joint BSC-CRG-IRB Research Program in Computational Biology
Barcelona Supercomputing Centre
Jordi Girona 29, 08034 Barcelona (Spain)

[c] Prof. V. Guallar
ICREA
Passeig Lluís Companys 23, 08010 Barcelona (Spain)

Supporting information for this article can be found under: <http://dx.doi.org/10.1002/cplu.201700030>. It contains details of the construction of the UNIK variants. Laccase activity assay: standard assay; microtiter plate assay. Kinetic parameters. Enzyme functionalisation. Mass spectrometry. Photoreduction experiments: light sources, filters and settings; effect of light on laccase activity; photoreduction kinetics; evolution of *p*-styrenesulfonate oxidation products. Computational details: homology modelling and system preparation; covalent docking of complex 1 and docking of [Ru(bpy)₃]²⁺; molecular dynamics and metadynamics; electronic coupling calculations.

This article is part of the "Biofest" Special Issue. To view the complete issue, visit: <https://doi.org/10.1002/cplu.v82.4>.

investigated bimolecular systems by associating the enzyme with palladium complexes (for alcohol oxidation) or ruthenium–polypyridyl photosensitisers (for epoxidation) taking advantage of the driving force of dioxygen reduction (1.23 V at pH 0).^[17,18] However, bimolecular systems are intrinsically limited by diffusion. Today, the well-known topography of their redox centres (there are tens of 3D structures available in databases) makes laccases perfect candidates for surface-targeted modulation of the enzyme function.

Most tagging procedures are generally unspecific as they are based on a general chemical character (e.g., nucleophilicity) of the side chains of surface amino acids. Amino acids with a reactive function are either naturally abundant at the protein surface (e.g., lysine, K) or might be involved in post-translational modifications (e.g., cysteines). Interestingly, there are only two lysine residues naturally present in the mature sequence of the laccase LAC3 from the fungus *Trametes* sp. C30.^[19] This natural paucity in lysine residues suggests a facile engineering of the enzyme to obtain variants with a unique lysine residue located at a desired position on the protein surface (UNIK variants). We report here on the site-directed reductive alkylation of both LAC3 and two of its UNIK variants with a custom ruthenium–polypyridyl-aldehyde complex, allowing the enzyme photoreduction to be probed from distinct surface locations relative to the redox-centres. Performed alongside experimental measurements, molecular simulations, which have proved recently to be effective in modelling laccase–substrate binding and reactivity,^[20,21] help us understand the structure and ET behaviour of such hybrid systems.

Results and Discussion

Design and expression of UNIK variants

LAC3 is a typical fungal laccase,^[19] recombinant forms of which can be easily expressed either in *Saccharomyces cerevisiae* for screening purposes or in the filamentous fungus *Aspergillus niger* for gram-per-litre-scale production.^[22,23] We have already challenged the plasticity of this enzyme,^[24,25] as well as its robustness in chemo-enzymatic catalysis experiments involving an ET either from a transition-metal catalyst or from photosensitisers.^[17,18,26,27] The *lac3* cDNA (GenBank AAR00925.1) encodes for 501 residues, among which are lysines K40 and K71 (LAC3 numbering).^[19] Therefore, lysine variants of LAC3 (designated UNIK) with a unique reactive amino group precisely located at the enzyme surface can be obtained through the replacement of any codon encoding a surface-located residue by an AAA/G codon (encoding a K residue) in a library of *lac3* sequences representing the natural variations found at positions 40 and 71. Because functional variants can be easily selected upon expression in yeast,^[24] such a surface-specific targeting strategy is potentially a powerful tool to probe the ET between the enzyme surface and its interior (i.e., the metal centres). To illustrate this approach, we prepared UNIK variants graftable in the vicinity of the T1 Cu^{II} site, at positions almost diametrically opposed to positions K40 and K71 originally found at the surface of LAC3 (Figure 1). From an examination of the natural varia-

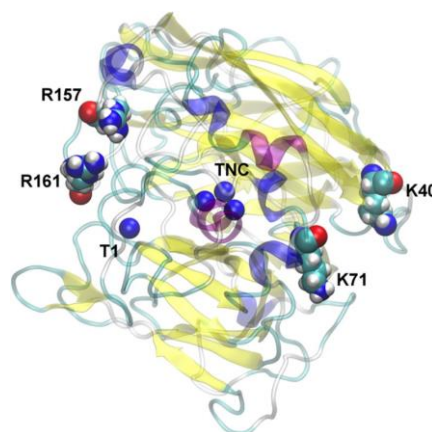


Figure 1. Structural model (ribbon) of LAC3 from *Trametes* sp. C30. Cu^{II} ions are depicted as blue spheres; amino acids subjected to mutagenesis are labelled (van der Waals volumes, CPK colours).

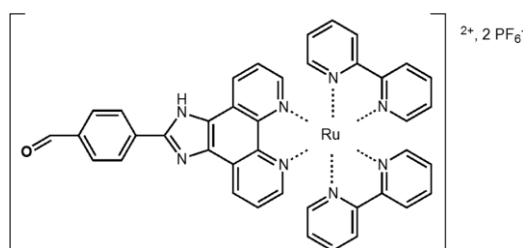
tion of the amino acid in loops surrounding the T1 Cu^{II} binding site in laccase sequences from the databases, we initially selected R157 and R161 as residues to substitute for a lysine group. Indeed, the variability at these two positions is high in fungal laccases and residues frequently found include lysines. The construction of UNIK variants included three steps of mutagenesis resulting in successive site-directed replacements of lysine codons 40 and 71 by codons corresponding to the most frequent amino acids found at these positions (M, Q or I at position 40 and H, A, R or E at position 71), followed by the replacement of codons 157 or 161 by a lysine codon. After each step, variants heterologously expressed in *S. cerevisiae* were screened for laccase activity on the substrate 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS). The second step yielded enzymes devoid of lysine (NoK); the third step, yielded UNIK₁₅₇ and UNIK₁₆₁ enzymes (see the Supporting Information for construction details). Among the 12 variants obtained for each construction [combinatorial in variations considered for positions 40 (three different residues) and 71 (four different residues)], we selected the UNIK₁₅₇ (M40, H71) and UNIK₁₆₁ (M40, H71) enzymes on the basis of their ABTS oxidation activity relative to that of the parental enzyme LAC3 (see the Supporting Information). Selected enzymes were further produced on a gram-per-litre scale by fermentation of recombinant *A. niger* strains, as previously described.^[23]

Functionalisation of LAC3 and UNIK variants

Primary amino groups present at the surface of proteins are reactive with aldehyde-containing molecules in a reductive alkylation reaction. The UNIK₁₆₁ (M40, H71) variant has been recently successfully derivatised with a 2,2,2-trifluoro-*N*-(4-oxo-butyl)-*N*-(1-pyren-4-ylmethyl-1*H*-[1,2,3]triazol-4-ylmethyl)acetamide moiety allowing a precise orientation of the enzyme at a multi-walled carbon nanotube electrode surface.^[28] To probe surface-accessible lysine residues of both UNIK and LAC3 enzymes we

selected ruthenium(II)–polypyridine-type chromophores for the following reasons: 1) their intense absorption band in the visible spectrum at $\lambda_{\text{max}} \approx 450$ nm (metal-to-ligand charge transfer, $\epsilon = 14500 \text{ M}^{-1} \text{ cm}^{-1}$).^[29] This band is not overlapping the $\lambda_{\text{max}} \approx 605$ nm band characteristics of the T1 Cu^{II} site of laccases (ligand-to-metal charge transfer, $\epsilon = 5600 \text{ M}^{-1} \text{ cm}^{-1}$)^[14] and therefore the 450/605 ratio can be used to evaluate the grafting efficiency; 2) their fluorescence properties ($\lambda_{\text{ex}} \approx 450$ nm, $\lambda_{\text{em}} \approx 600$ nm),^[13] which are useful to visualise grafted enzymes on SDS-PAGE gels and grafted peptides after digestion and HPLC separation; and 3) their capability to photoreduce laccases upon visible light excitation in the presence of an electron donor.^[26]

We used a heteroleptic ruthenium(II) complex with one of the bipyridine ligands containing a fused imidazole motif and a benzaldehyde (complex 1, Scheme 1). This complex is similar to the $[\text{Ru}(\text{bpy})_3]^{2+}$ ($\text{bpy} = 2,2'$ -bipyridine) complex we have previously used in bimolecular reactions.^[26] LAC3, UNIK₁₅₇ and UNIK₁₆₁ were functionalised with 1 by reductive alkylation in mild conditions (see the Supporting Information).^[30]



Scheme 1. The structure of complex 1, $[(\text{bpy})_2\text{Ru}(\text{fimbzl})][\text{PF}_6]_2 \cdot 2\text{H}_2\text{O}$ ($\text{fimbzl} = 4$ - $[(1,10)\text{phenanthroline}[5,6-d]\text{imidazol-2-yl}]$ benzaldehyde).

Representative UV/Vis spectra of enzyme–1 hybrids (1–LAC3 and 1–UNIK₁₅₇) recorded after a gel filtration, are presented in Figure 2. Covalent linkage of complex 1 to the enzymes was addressed by subjecting the grafted laccases to electrophoresis in denaturing conditions. Under UV irradiation (360 nm) of the gel, the fluorescence of the ruthenium is coincident to the laccase band in samples subjected to reductive alkylation in the presence of complex 1 (Figure 3). At the pH of the reaction (7.4), lysine groups as well as the amino N terminus can be

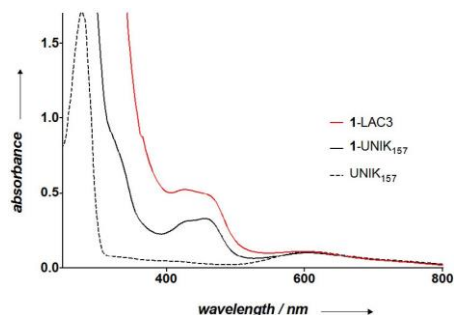


Figure 2. Representative UV/Vis spectra of bare and 1-grafted enzymes. Enzyme concentration ca. $16 \mu\text{M}$ in Britton–Robinson buffer (pH 5).

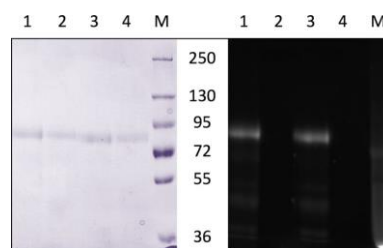


Figure 3. Representative separation of grafted enzymes on a 7% SDS-PAGE gel. Left panel: white-light irradiation (Coomassie Blue staining); right panel: UV irradiation (360 nm). M: molecular weight markers. Apparent masses are expressed in kDa. Lane 1: 1–UNIK₁₆₁ (5 μg); lane 2: UNIK₁₆₁ (5 μg); lane 3: 1–LAC3 (5 μg); lane 4: LAC3 (5 μg).

subjected to reductive alkylation. Batch after batch, the grafting procedure yielded, respectively, one and two complex 1 moieties per UNIK and LAC3 enzymes (based on the A_{450}/A_{605} ratio, Figure 2 and Table 1).

The complex 1 to laccase ratios were further confirmed by analysing the metal content of the grafted enzymes by inductively coupled plasma mass spectrometry (ICP–MS, Table 1). Eventually, the location of the graft at the surface of enzymes was investigated by Edman degradation and MALDI–TOF MS analysis. The sequencing of each enzyme revealed that the amino terminus of the hybrids was an alanine, the first residue of the mature enzyme, indicating that the N terminus was not blocked by complex 1. Trypsin digestion of UNIK₁₆₁ modified

Table 1. Spectroscopic features, sequence data and metal content of the 1-laccase hybrids.					
Hybrid	A_{450}/A_{605}	Amino terminus	Metal content Ru/Cu ^[a]	1-labelled tryptic peptide ^[b]	Theoretical mass of peptide [Da] ^[c]
1–LAC3	4.73	A	1.8–2.2/4	AAVMNDQFPGPLIAGNKGNDFQINVIDNLSNSTMSTTIHWHGFFQKGT-NWADGAFAFVNQCPSIAGNSFLYDFATDQAGTFWYHSHLSTQYCDGLR	12342.31
1–UNIK ₁₅₇	2.97	A	1.1/4	GPMVVYDPDPPHASLYDVDDSDVITLSWDVYHTAAKLGAR	5129.944
1–UNIK ₁₆₁	3.22	A	1.2/4	LGAKFPAGADTLINGLGR	2578.28 (2580.20)

[a] Measured by ICP–MS. [b] The functionalisation by complex 1 apparently prevents the cleavage of the peptide bond immediately after the lysine residue. [c] m/z is given in brackets.

by complex **1** yielded a single fluorescent peptide corresponding to a singly charged ion of m/z 2580.20 in the mass spectrum. Both an increment of 722.14 Da in mass (i.e., 738.14–16 Da corresponding to a loss of an oxygen atom in the reductive alkylation reaction) and the characteristic isotopic distribution of ruthenium unambiguously identified K161 as the amino acid grafted with complex **1** in the tryptic peptide of the corresponding 1–UNIK enzyme (Table 1 and see the Supporting Information for details). For LAC3, the theoretical mass of the fluorescent peptide is 12342.31 Da including a mass increment of 1440.28 Da corresponding to two molecules of complex **1**. Probably because of the mass of that peptide, data collected so far on the tryptic digestion of LAC3 are unfortunately inconclusive. However, the results of the Edman degradation, the metal quantification and the UV/Vis spectra support a double grafting to an enzyme surface that contains only two lysine residues at positions 40 and 71. In summary, under our experimental conditions, the reductive alkylation reaction appears selective for the primary amine group of surface-accessible lysine residues. Therefore, the reductive alkylation of UNIK variants warrants a single functionalisation of the laccase surface.

Kinetic characterisation of bare and ruthenium-grafted laccase variants

The most selective surface functionalisation process would be ineffective if it abolishes the original enzyme function. Laccase is a particularly robust enzyme and experimental conditions used for the reductive alkylation are considered mild compared to other chemical procedures. Nevertheless, the grafting procedure is potentially non-innocent towards enzymes function. Furthermore, the presence of the graft itself at the

enzyme surface might interfere with the redox properties of laccases.

The influence of the grafting procedure and/or the presence of the graft on laccase reactivity was evaluated by performing an analysis of the kinetic properties of the bare and ruthenium-grafted laccases. Oxidation kinetic parameters obtained in the presence of ABTS are reported in Table 2 (see Table S2 for data obtained with syringaldazine). First, combinations of the mutations found in the two UNIK variants (i.e., at positions 40, 71 and 157 or 161) have apparently no major effect on the catalytic efficiency of the variant enzymes as compared to the parental enzyme LAC3 (Table 1, entries 1, 3 and 5). This is consistent with the results of the screening based on ABTS oxidation used to select these variants. Second, the grafting procedure does not compromise the catalytic efficiency of the enzymes. Thus, the reductive alkylation appears to be a valuable method to functionalise the surface of laccases. Third, upon grafting, complex **1** does not significantly interfere with the catalytic efficiencies of the variants. From a study on a bimolecular photocatalytic system made of LAC3 and a ruthenium–polypyridyl complex similar to complex **1** (having a benzoic acid moiety instead of a benzaldehyde), we know that such complexes in solution and in excess can interfere with the oxidation activity of the laccase on the natural substrates.^[26] Apparently, grafting helps circumvent this issue by setting both the complex/enzyme ratio and the location of the complex relative to the substrate oxidation site. In 1–LAC3, the grafted positions K40 and K71 are almost diametrically opposed to the T1 Cu^{II} site (Cu^{II}–Ru^{II} distances are between 37 and 45 Å, as estimated from molecular simulations, see Table 3). The kinetics obtained with 1–LAC3 are consistent with grafts that cannot physically interfere with the substrate oxidation activity of the enzyme. On the contrary, in 1–UNIK hybrids, complex **1** is grafted in the vicinity of the T1 Cu^{II} site (Figure 1). Therefore, the absence of significant variation in the catalytic efficiencies of grafted variants underlines both the plasticity of the enzyme surface and the overall robustness of the enzyme.

In order to understand the molecular basis of the role of the grafted complex **1** in native catalysis, models of 1–laccase were built. Reductive alkylation between laccase variants and complex **1** was simulated with covalent docking.^[31] A 50 ns molecular dynamics simulation was run to refine the binding mode and to take into account the interaction with explicit solvent molecules. Molecular models for 1–UNIK₁₅₇ and 1–UNIK₁₆₁ are presented in Figure 4 (molecular dynamics RMSD plots of both the protein backbone and complex **1** are illustrated in Fig-

Table 2. Apparent kinetic parameters for laccase variants on ABTS.^[a]

Entry	Enzyme	k_{cat} [s ⁻¹]	K_M [M]	k_{cat}/K_M [s ⁻¹ M ⁻¹]
1	LAC3	246 ± 1	3.72 × 10 ⁻⁴ ± 10	6.62 × 10 ⁵
2	1–LAC3	173 ± 7	3.42 × 10 ⁻⁴ ± 44	5.10 × 10 ⁵
3	UNIK ₁₅₇	247 ± 1	4.07 × 10 ⁻⁴ ± 12	6.06 × 10 ⁵
4	1–UNIK ₁₅₇	188 ± 2	4.61 × 10 ⁻⁴ ± 23	4.07 × 10 ⁵
5	UNIK ₁₆₁	204 ± 3	3.08 × 10 ⁻⁴ ± 24	6.63 × 10 ⁵
6	1–UNIK ₁₆₁	195 ± 1	2.92 × 10 ⁻⁴ ± 6	6.70 × 10 ⁵

[a] In Britton–Robinson buffer, pH 5.5, 25 °C; enzyme concentration: ca. 5 × 10⁻¹⁰ M; ABTS concentration: 0.1–10 × 10⁻³ M.

Table 3. Electronic coupling (EC) and distances (d) between Ru and Cu atoms.

Cu ^{II} ion	1–LAC3		1–UNIK ₁₆₁		1–UNIK ₁₅₇		1 + LAC3			
	K40	K71	K40	K71	K40	K71	K40	K71		
	EC [eV]	d [Å]	EC [eV]	d [Å]	EC [eV]	d [Å]	EC [eV]	d [Å]		
T1	2 × 10 ⁻¹³	44.9	6 × 10 ⁻¹¹	36.8	1 × 10 ⁻⁷	23.5	1 × 10 ⁻⁷	19.8	3 × 10 ⁻⁴	11.2
T2	4 × 10 ⁻¹¹	30.7	2 × 10 ⁻⁹	26.4	5 × 10 ⁻⁹	25.5	6 × 10 ⁻¹⁰	23.6	3 × 10 ⁻⁸	24.8
T3 α	2 × 10 ⁻¹⁰	32.6	3 × 10 ⁻⁹	32.0	2 × 10 ⁻⁹	25.8	3 × 10 ⁻⁹	23.3	3 × 10 ⁻⁷	22.1
T3 β	2 × 10 ⁻¹²	34.5	1 × 10 ⁻⁹	28.8	5 × 10 ⁻⁸	20.5	1 × 10 ⁻⁷	19.3	8 × 10 ⁻⁸	21.8

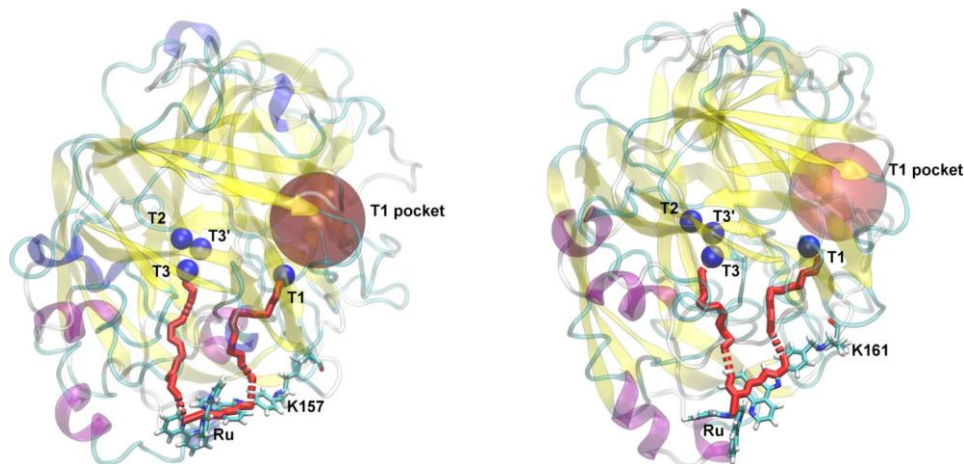


Figure 4. Position of complex 1 in 1-UNIK₁₅₇ (left) and 1-UNIK₁₆₁ (right) obtained from covalent docking plus molecular dynamics simulations. Potential ET pathways are depicted in red. The surface area just above the mononuclear T1 Cu centre is highlighted as a red sphere.

ure S1). Our models suggest that complex 1 is located at the enzyme surface far from the T1 Cu^{II} site entrance (the Ru^{II}–Cu^{II} distance is > 19 Å, see below in Table 3), and likely incapable of hindering the approach of the substrate (Figure 4). Due to this distant location, the extra positive charge added after reductive alkylation (taking into account two positive charges for complex 1) should not alter the oxidation process, in agreement with the kinetic data (Table 2). In the 1-LAC3 model, contributions from positions K40 and K71 were analysed separately for simplicity (Figure S3). In both cases, complex 1 was initially predicted as being partially buried in the cavity adjacent to the TNC centre. In order to better establish the exposed/buried nature of complex 1, we used an advanced metadynamics sampling technique (see the Supporting Information for details).^[34] For both grafting positions, we found that complex 1 is more stable outside the cavity and exposed to the bulk solvent (Figure S2). In summary, the grafting proposed does not compromise the natural catalytic properties of the enzymes. Therefore, LAC3 and UNIK variants can be used to probe the orientation of a graft relative to the catalytic centres.

Photoreduction

As illustrated by several studies, grafting the surface of a metalloenzyme with ruthenium–polypyridyl complexes is a powerful way to study the transfer of electrons to metal centres embedded in protein matrices.^[32] By using a bimolecular system we have recently demonstrated that LAC3 can be reduced by excited photosensitisers in the presence of a sacrificial electron donor. In the first instance, 1-LAC3 and 1-UNIK hybrids were subjected to photoreduction experiments, as previously described.^[26,27] In dioxygen-deprived solutions, in the presence of EDTA as electron donor and under constant white light irradiation, all 1-grafted enzymes were progressively reduced as a function of time (Figure 5). After one hour of irradiation in

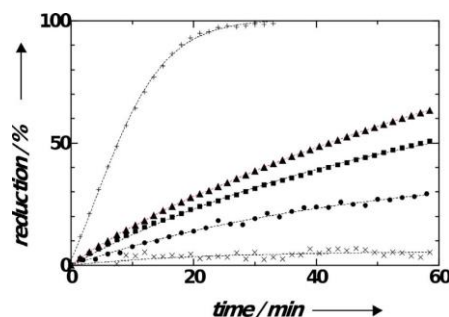


Figure 5. Photoreduction of laccase–ruthenium hybrids. Argon-flushed Britton–Robinson buffer solutions (pH 5) containing 3.0×10^{-5} M hybrids in the presence of 1.5 mM EDTA (electron donor). Irradiation was performed with a Dolan-Jenner MI-150 illuminator (Edmund) equipped with a 150 W EKE Quartz Halogen lamp filtered with a 450 nm OD 2 longpass filter (Techspec) and guided with an optic fibre (power density of 235 mW cm^{-2} prior to filtration). LAC3 enzyme (X); 1-LAC3 hybrid (λ); 1-UNIK₁₆₁ hybrid (v); 1-UNIK₁₅₇ hybrid (σ); LAC3 enzyme + [Ru(bpy)₃]²⁺ bimolecular system (+). See the Supporting Information for the fitted curves. Data points represent an average of three independent measurements.

a degassed solution, the extend of reduction (monitored as a time-dependent decrease of the T1 MLCT) of 1-LAC3, 1-UNIK₁₆₁ and 1-UNIK₁₅₇ was found to be, respectively approximately 20%, 40% and 50% on top of the observed photoreduction of the bare enzyme (less than 10% in a control experiment, Figure 5).

Reduction kinetics of the two 1-UNIK hybrids appear similar to each other and are distinguishable from that of the hybrid 1-LAC3. UNIK hybrids differ from LAC3 in both the number of surface grafts (two in 1-LAC3 and one in 1-UNIK hybrids) and in terms of the area over which complex 1 is grafted (vide supra). From the location of the graft at the enzyme surface, the photoreduction of the hybrids potentially arises from three distinct phenomena: 1) a direct reduction of the T1 Cu^{II} site by an

excited complex **1** (**1***); 2) a reduction of the T1 Cu^{II} site consecutive to the initial reduction of the TNC by **1*** (reverse intramolecular reaction); and 3) a photo-induced ET between an excited complex **1** moiety of one hybrid and the T1 Cu^{II} site of another (bimolecular reaction). The latter phenomenon, observable for example, in concentrated Ru–P450cam hybrid solutions,^[33] is likely to be minor here considering that no significant variation of the reduction rates was observed over a large range of hybrid concentrations (i.e., from 4 to 9.0×10^{-5} M). Unless considering a long-range ET within the protein matrix, the first proposal seems difficult to apply to the hybrid **1**–LAC3, leaving potentially the TNC as the primary electron acceptor. On the contrary, in **1**–UNIK hybrids, the location of complex **1** in the vicinity of the T1 Cu^{II} favours a direct reduction of this centre.

Modelling was again used to address these suppositions on a molecular basis. Ten frames, equispaced in time over the last 10 ns of molecular dynamics runs, were taken and both the electronic coupling (EC) and distances between the Ru^{II} and Cu^{II} ions were calculated (Table 3). In agreement with the photoreduction experiment, calculated EC values are similar in the two UNIK variants. As expected, the T1 Cu^{II} is probably the first electron acceptor for both hybrids, although it should be noted that **1*** might also transfer electrons to one of the T3 Cu^{II} ions, at least in **1**–UNIK₁₅₇ (see EC values in Table 3). In the models presented in Figure 4, complex **1**, bound either to K157 or K161, is located in a shallow depression of the enzyme's surface such that the Ru^{II} ion is almost equidistant to both the T1 and the T3 Cu^{II} ions.^[37] Then, upon irradiation, ET occurs between metal centres 20–25 Å away from each other (long-range ET). In **1**–LAC3, the distances obtained from molecular models suggest that **1*** might potentially transfer electrons most favourably to the TNC and with a relative higher efficiency from position 71 (Table 3). Here, the metal-to-metal distance (Ru^{II}–Cu^{II} $d \approx 30$ Å) is even higher than in **1**–UNIK hybrids and this correlates with the slower kinetics of photoreduction (Figure 5).

Overall, the kinetics of photoreduction measured for **1**-grafted enzymes are much slower than those obtained with the bimolecular system consisting of LAC3 and a [Ru(bpy)₃]²⁺ complex (Figure 5). In order to compare EC between these two configurations, the bimolecular system was simulated by non-covalently docking a [Ru(bpy)₃]²⁺ complex into the T1 Cu^{II} cavity, followed by electronic coupling calculations. The most favourable location of the [Ru(bpy)₃]²⁺ complex at the enzyme surface (Figure 6) is similar to that earlier obtained by Monte Carlo simulations by Kurzev et al. for a cyclometalated complex of ruthenium used as a laccase substrate.^[35] In the model presented in Figure 6, the Ru^{II}-to-T1 Cu^{II} distance is half that calculated for grafted UNIKs (Table 3). Furthermore, EC in the bimolecular system is three orders of magnitude larger than in grafted systems, which is consistent with its faster T1 Cu^{II} reduction. These calculations suggest that a faster T1 Cu^{II} reduction should be obtained by any design that reduces the Ru^{II}-to-T1 Cu^{II} distance in UNIK variants grafted with a ruthenium complex, for example, a graft bearing a shorter bipyridine substituent and/or varying the grafting point at the enzyme sur-

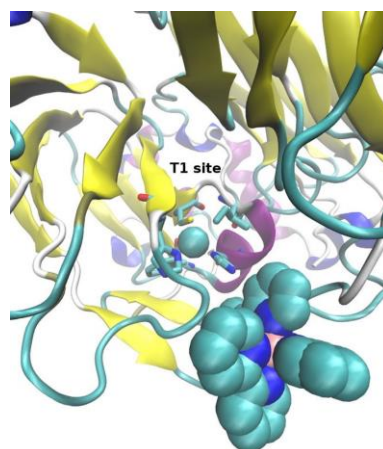


Figure 6. Position of a [Ru(bpy)₃]²⁺ complex in front of the T1 copper centre obtained from non-covalent docking.

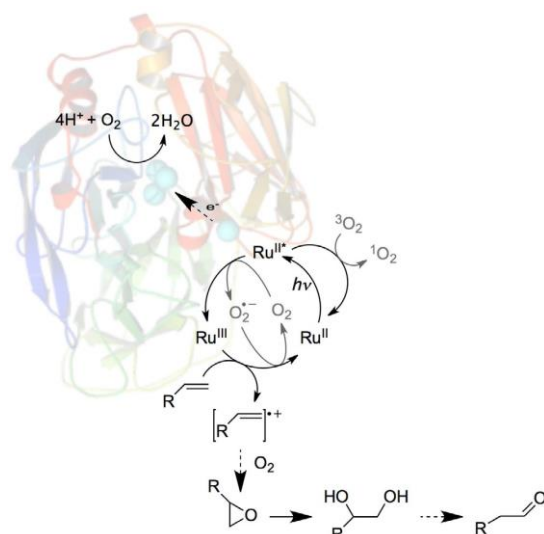
face. On the other hand, it is interesting to compare the modest efficiency of the T1 Cu^{II} photoreduction observed with the **1**–UNIK hybrid with the large benefit obtained from a pyrene–UNIK₁₆₁ hybrid in the operation of a laccase biocathode.^[28] Both systems comply with a standard functioning of the enzyme, that is, a cumulative one-electron reduction of the four Cu^{II} ions. Strategies for the functionalisation of the UNIK₁₆₁ variant being identical, differences in performance probably arise from the contact area between the material delivering electrons (carbon nanotube surface vs. a ruthenium–polypyridyl complex) and the enzyme surface. This strongly supports that a faster T1 Cu^{II} photoreduction can be obtained by varying the grafting point in **1**–UNIK hybrids.

Photocatalysis

We have recently described the first photo-oxidation of olefins coupled to the light-driven reduction of O₂ by a [Ru(bpy)₃]²⁺–laccase bimolecular system.^[18] With such a minimal system, olefins oxidation probably occurs through a radical mechanism in which O₂ acts both as a renewable electron acceptor and as O atom donor (Scheme 2).

To complete the present characterisation of **1**-grafted enzymes, the ability of **1**–LAC3 and **1**–UNIK enzymes to oxidise *p*-styrenesulfonate was evaluated. For simplicity, of the two UNIK enzymes, only **1**–UNIK₁₆₁ was evaluated as it appeared from calculations that complex **1** might be electronically coupled equally to both the Cu^{II} T1 and the TNC centres in **1**–UNIK₁₅₇ (Table 3). Upon irradiation, **1**-grafted systems catalysed the formation of *p*-styrenesulfonate oxidised species as monitored by ¹H NMR spectroscopy (Table 4).

Substantial amounts of oxidation products (the epoxide and the corresponding diol and aldehyde) were detected in the aerated mixture of the **1**–UNIK₁₆₁ system.^[38] By contrast, photo-oxidation of *p*-styrenesulfonate in the presence of **1**–LAC3 yielded only approximately one-sixth of the amount of prod-



Scheme 2. Photo-oxidation of olefins coupled to the light-driven reduction of O_2 by a Ru-polypyridyl complex-laccase.^[18] The paths depicted in grey do not lead to product formation in the absence of enzyme. Ribbon model of laccase: Cu ions are depicted as cyan spheres.

System	Products $\times 10^{-6}$ M			Total $\times 10^{-6}$ M
1 + LAC3	145 \pm 25	140 \pm 15	55 \pm 25	340
1 -LAC3	n.d.	n.d.	70 \pm 5	70
1 -UNIK ₁₆₁	150 \pm 20	160 \pm 27	80 \pm 10	390

[a] Conditions: 24 h irradiation with filtered white LED light (ca. 5 mW cm⁻², 450 nm < λ < 700 nm); [1] = [LAC3] = [1-LAC3] = [1-UNIK₁₆₁] = 30 $\times 10^{-6}$ M relative to the enzyme, [*p*-styrenesulfonate] = 10 $\times 10^{-3}$ M; R = phenyl sulfonate; deuterated phosphate buffer, pH 6.0, T = 25 °C. n.d.: not detected; in the absence of light, enzyme or dioxygen, no products were detected. Mean of three independent experiments.

ucts obtained with **1**-UNIK₁₆₁ (Table 4). Therefore, with two units of complex **1** grafted to 180° from the T1 centre the **1**-LAC3 system is poorly active as a photocatalyst. Altogether, the behaviour of the two hybrid systems as photocatalysts correlates well with Ru^{II}-Cu^{II} EC values calculated for **1**-LAC3 that are consistently lower than those calculated for **1**-UNIK₁₆₁ (Table 3).

As for the bimolecular system (Table 4, entry 1), use of the **1**-UNIK₁₆₁ system led to the accumulation of a *primo* formed epoxide in a reaction previously shown to be limited by O_2 .^[18] The initial formation of an enyl radical cation is likely followed by the insertion of one O atom from O_2 into the electron-deficient *p*-styrenesulfonate radical (Scheme 2).^[18] According to this reaction scheme, 3.90 $\times 10^{-4}$ M of products obtained here with **1**-UNIK₁₆₁ necessitated a minimum of 3.00 $\times 10^{-4}$ M of O_2 from which one-third accounts for a four-electron reduction

into water and two-thirds for a subsequent O atom insertion. This concentration of O_2 is slightly greater than that of the dioxygen dissolved in the solution before irradiation (ca. 2.50 $\times 10^{-4}$ M) suggesting that some O_2 was obtained after dissolution from the gas phase. So, under photocatalytic conditions the grafted system appears susceptible to follow a regime similar to that of the bimolecular one (Table 4, entry 1).^[18] At this point, both photoreduction experiments and calculated EC values suggest that **1**-UNIK₁₆₁ and **1**-UNIK₁₅₇ constructions are suboptimal. However, setting both the complex/enzyme ratio and the location of the complex at the enzyme surface, we readily addressed the major issues intrinsic to the bimolecular system: 1) the need to maintain a relatively high concentration of both partners to favour intermolecular ET reactions; 2) the dependence on a high sensitizer-to-laccase ratio to improve yields; and 3) the rapid inhibition of the activity of the laccase for the oxidation of the natural substrate at relatively low sensitizer concentration.^[18,26] Therefore, building on the concept of UNIK enzymes, it is worth investing in the development of new grafts and/or in the design of new variants allowing to decrease the Ru^{II}-to-T1 Cu^{II} distance in order to improve the efficiency of the photocatalyst. Such developments are largely more promising than developing the bimolecular system.

Conclusion

The lysine-free laccase we created in this study allows the design of single surface-located lysine variants, designated UNIKs, and the precise engineering of the enzyme surface by post-functionalisation. The functionalisation of different areas of the enzyme surface is easily obtained by reductive alkylation without compromising the enzyme core function, that is, the intramolecular ET and dioxygen reduction. Importantly, UNIKs are produced on a gram-per-litre scale by fermentation of recombinant *A. niger* strains, ensuring a comfortable physico-chemical characterisation of the grafted enzymes, as well as a non-compromised future use of these enzymes as biocatalysts. Using a ruthenium photo-sensitizer, we demonstrate that this strategy offers valuable tools to probe the ET between the enzyme surface and both the T1 and the TNC copper centres located in its interior. The computed Ru^{II}-to-Cu^{II} centre distances correlate well with the photoreducibility and in turn with the reactivity of the grafted enzymes. Overall, results obtained from this study allow a better understanding of the system from a fundamental point of view. From this experimental ground, we believe that molecular simulations can guide the design of improved hybrid biocatalysts as well as the precise orientation of the enzyme at a material surface to obtain smart materials.

Experimental Section

The [Ru(bpy)₃]²⁺ complex was purchased from Sigma-Aldrich. [(bpy)₂Ru(fimbzl)][(PF₆)₂·2H₂O] [where fimbzl = 4-((1,10)phenanthroline[5,6-*d*]imidazol-2-yl)benzaldehyde] was synthesised according to the procedure described by Gholamkhas and co-workers.^[36] Oxygen consumption was measured by polarography using

a model 781 oxygen meter (Strathkelvin Instruments, Motherwell, UK) with a micro Clark electrode fitted to a temperature-controlled glass chamber. Irradiation of the sample was performed through the glass chamber. Catalytic tests were performed in air on the bench in 2 mL vials.

Acknowledgements

L.S. was the recipient of a Ministère de l'Éducation Nationale fellowship. This study was supported by grants from the Agence Nationale de la Recherche (ANR-09-BLANC-0176 and ANR-15-CE07-0021-01) and from the Ministerio de Economía, Industria y Competitividad (CTQ2016-79138-R). We thank Elise Courvoisier-Dezord from the Plateforme AVB (AMU): Analyse et Valorisation de la Biodiversité and Yolande Charmasson for help in the production of the recombinant enzymes, as well as Pascal Mansuelle and Régine Lebrun from the Plateforme Protéomique (CNRS-AMU) for help in acquiring mass spectrometry data.

Conflict of interest

The authors declare no conflict of interest.

Keywords: electron transfer · enzyme models · multi-copper oxidase surfaces · photosensitisers · targeted functionalisation

- [1] C. A. Denard, H. Ren, H. Zhao, *Curr. Opin. Chem. Biol.* **2015**, *25*, 55–64.
- [2] W. H. Brondyk, *Methods Enzymol.* **2009**, *463*, 131–147.
- [3] C. Elena, P. Ravasi, M. E. Castelli, S. Peiru, H. G. Menzella, *Front. Microbiol.* **2014**, *5*, 21.
- [4] M. G. Roig, J. F. Kennedy, *Critic. Rev. Biotech.* **1992**, *12*, 391–412.
- [5] J. Eyzaguirre, *Biol. Res.* **1996**, *29*, 1–11.
- [6] G. DeSantis, J. B. Jones, *Curr. Opin. Biotechnol.* **1999**, *10*, 324–330.
- [7] B. G. Davis, *Curr. Opin. Biotechnol.* **2003**, *14*, 379–386.
- [8] T. Montes, V. Grazu, F. Lopez-Gallego, J. A. Hermoso, J. M. Guisan, R. Fernandez-Lafuente, *Biomacromolecules* **2006**, *7*, 3052–3058.
- [9] V. Stepankova, S. Bidmanova, T. Koudelakova, Z. Prokop, R. Chaloupkova, J. Damborsky, *ACS Catal.* **2013**, *3*, 2823–2836.
- [10] J. E. Moses, A. D. Moorhouse, *Chem. Soc. Rev.* **2007**, *36*, 1249–1262.
- [11] C. D. Spicer, B. G. Davis, *Nat. Commun.* **2014**, *5*, 4740.
- [12] T. Tron in *Encyclopedia of Metalloproteins* (Eds.: R. H. Kretsinger, V. N. Uversky, E. A. Permyakov), Springer, New York, **2013**, pp. 1066–1070.
- [13] D. M. Mate, M. Alcalde, *Microb. Biotechnol.* **2016**, DOI: 10.1111/1751-7915.12422.
- [14] E. I. Solomon, M. J. Baldwin, M. D. Lowery, *Chem. Rev.* **1992**, *92*, 521–542.
- [15] A. J. Augustine, C. Kjaergaard, M. Qayyum, L. Ziegler, D. J. Kosman, K. O. Hodgson, B. Hedman, E. I. Solomon, *J. Am. Chem. Soc.* **2010**, *132*, 6057–6067.
- [16] O. V. Morozova, G. P. Shumakovich, S. V. Shlevy, Ya. I. Yaropolov, *Appl. Biochem. Microbiol.* **2007**, *43*, 523–535.
- [17] Y. Mekmouche, L. Schneider, P. Rousselot-Pailley, B. Faure, A. J. Simaan, C. Bochet, M. Reglier, T. Tron, *Chem. Sci.* **2015**, *6*, 1247–1251.
- [18] L. Schneider, Y. Mekmouche, P. Rousselot-Pailley, A. J. Simaan, V. Robert, M. Reglier, A. Aukauloo, T. Tron, *ChemSusChem* **2015**, *8*, 3048–3051.
- [19] A. Klonowska, C. Gaudin, M. Asso, A. Fournel, M. Reglier, T. Tron, *Enz. Mic. Tech.* **2005**, *36*, 34–41.
- [20] E. Monza, M. F. Lucas, S. Camarero, L. C. Alejandre, A. T. Martínez, V. Guallar, *J. Phys. Chem. Lett.* **2015**, *6*, 1447–1453.
- [21] I. Pardo, G. Santiago, P. Gentili, F. Lucas, E. Monza, F. J. Medrano, C. Galli, A. T. Martínez, V. Guallar, S. Camarero, *Catal. Sci. Technol.* **2016**, *6*, 3900–3910.
- [22] V. Robert, Y. Mekmouche, P. R. Pailley, T. Tron, *Curr. Genomics* **2011**, *12*, 123–129.
- [23] Y. Mekmouche, S. Zhou, A. M. Cusano, E. Record, A. Lomascolo, V. Robert, A. J. Simaan, P. Rousselot-Pailley, S. Ullah, F. Chaspoul, T. Tron, *J. Biosci. Bioeng.* **2014**, *117*, 25–27.
- [24] A. M. Cusano, Y. Mekmouche, E. Meglec, T. Tron, *FEBS J.* **2009**, *276*, 5471–5480.
- [25] Y. Liu, A. M. Cusano, E. C. Wallace, Y. Mekmouche, S. Ullah, V. Robert, T. Tron, *Int. J. Biol. Macromol.* **2014**, *69*, 435–441.
- [26] A. J. Simaan, Y. Mekmouche, C. Herrero, P. Moreno, A. Aukauloo, J. A. Delaire, M. Reglier, T. Tron, *Chem. Eur. J.* **2011**, *17*, 11743–11746.
- [27] T. Lazarides, I. V. Sazanovich, A. J. Simaan, M. C. Kafentzi, M. Delor, Y. Mekmouche, B. Faure, M. Reglier, J. A. Weinstein, A. G. Coutsolelos, T. Tron, *J. Am. Chem. Soc.* **2013**, *135*, 3095–3103.
- [28] N. Lalaoui, P. Rousselot-Pailley, V. Robert, Y. Mekmouche, R. Villalonga, M. Holzinger, S. Cosnier, T. Tron, A. Le Goff, *ACS Catal.* **2016**, *6*, 1894–1900.
- [29] K. Kalyanasundaram, *Coord. Chem. Rev.* **1982**, *46*, 159–244.
- [30] J. M. McFarland, M. B. Francis, *J. Am. Chem. Soc.* **2005**, *127*, 13490–13491.
- [31] K. Zhu, K. W. Borrelli, J. R. Greenwood, T. Day, R. Abel, R. S. Farid, E. Harder, *J. Chem. Info. Model.* **2014**, *54*, 1932–1940.
- [32] a) C. Mayer, D. G. Gillingham, T. R. Ward, D. Hilvert, *Chem. Commun.* **2011**, *47*, 12068–12070; b) J. R. Winkler, H. B. Gray, *Chem. Rev.* **1992**, *92*, 369–379; c) O. Kokhan, N. S. Ponomarenko, P. R. Pokkuluri, M. Schiffer, K. L. Mulfort, D. M. Tiede, *J. Phys. Chem. B* **2015**, *119*, 7612–7624; d) M. E. Ener, Y.-T. Lee, J. R. Winkler, H. B. Gray, L. Cheruzel, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 18783–18786.
- [33] J. Contzen, S. Kostka, R. Kraft, C. Jung, *J. Inorg. Biochem.* **2002**, *91*, 607–617.
- [34] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562–12566.
- [35] S. A. Kurzeev, A. S. Vilesov, T. V. Fedorova, E. V. Stepanova, O. V. Koroleva, C. Bukh, M. J. Bjerrum, I. V. Kurnikov, A. D. Ryabov, *Biochemistry* **2009**, *48*, 4519–4527.
- [36] B. Gholamkhash, K. Koike, N. Negishi, H. Hori, T. Sano, K. Takeuchi, *Inorg. Chem.* **2003**, *42*, 2919–2932.
- [37] The T3 nomenclature follows that proposed in Augustine et al., Ref. [15].
- [38] At the pH of the reaction it is expected that the epoxide unambiguously spontaneously into a diol that can be itself subject to photo-oxidation into an aldehyde.

Manuscript received: January 25, 2017

Revised: February 2, 2017

Accepted Article published: February 3, 2017

Final Article published: March 2, 2017

CHEMPLUSCHEM

Supporting Information

Probing the Surface of a Laccase for Clues towards the Design of Chemo-Enzymatic Catalysts

Viviane Robert,^[a] Emanuele Monza,^[b] Lionel Tarrago,^[a] Ferran Sancho,^[b] Anna De Falco,^[a] Ludovic Schneider,^[a] Eloïne Npetgat Ngoutane,^[a] Yasmina Mekmouche,^[a] Pierre Rousselot Pailley,^[a] A. Jalila Simaan,^[a] Victor Guallar,^[b, c] and Thierry Tron^{*[a]}

cplu_201700030_sm_miscellaneous_information.pdf

Details of UNIK variants construction	2
Laccase activity assay	3
Standard assay	3
Micro-titer plate assay	3
Kinetic parameters	3
Enzymes functionalization	4
Mass spectrometry.....	4
Photo reduction experiments	4
Light sources, filters and settings.....	4
Effect of light on laccase activity.....	4
Photoreduction kinetics	5
Evolution of <i>p</i> -styrene sulfonate oxidation products	5
Computational details	5
Homology modelling and system preparation.....	5
Covalent Docking of complex 1 and docking of [Ru(bpy) ₃] ²⁺	5
Molecular dynamics and metadynamics.....	6
Electronic coupling calculations	7

* corresponding author;

E-mail address: thierry.tron@univ-amu.fr

Details of UNIK variants construction.

Substitutions of the natives K40, K71, R157 or R161 residues in the LAC3 sequence were obtained by site directed mutagenesis (QuikChange®, Stratagene™) of the *lac3* gene present in the plasmid pAK145.¹ The mutagenic oligonucleotides (obtained from Invitrogen™) are presented table S11.

Table S11. Sequence of mutagenic primers. In **bold**, codon variation relative to the *lac3* native sequence.

Primer name	Sequence (5'→3')	Step
K40M Fwd K40M Rev K40Q Fwd K40Q Rev K40I Fwd K40I Rev	GATCGCTGGAAAC AT GGGAGACAACTTC GAAGTTGTCTCC AT GTTTCCAGCGATC CGCTGGAAAC CCAG GGAGACAAC GTTGTCTCC CTGG TTCAGCGG GATCGCTGGAAAC ATC GGAGACAACTTC GAAGTTGTCTCC GAT GTTTCCAGCGATC	1
K71H Fwd K71H Rev K71A Fwd K71A Rev K71R Fwd K71R Rev K71E Fwd K71E Rev	CACGGCTTCTTCCAG CAC GGGCACTAATTGGGC GCCCAATTAGTGCC GTG CTGGAAGAAGCCGTG CACGGCTTCTTCCAG GCAG GCACTAATTGGGC GCCCAATTAGTGCC TGC CTGGAAGAAGCCGTG CGGCTTCTTCCAG AGAG GCACTAATTG CAATTAGTGCC TCT CTGGAAGAAGCCG CGGCTTCTTCCAG GAAG GCACTAATTG CAATTAGTGCC TTCT CTGGAAGAAGCCG	2
R157K Fwd R157K Rev	GTACCACACTGCCGCG AA ACTCGGCGCTCGTTTC GAAACGAGCGCCGAG TTT CGCGGCAGTGTGGTAC	3
R161K Fwd R161K Rev	CGTCTCGGCGCT AA ATTCCTGCTGGG CCCAGCAGGGA TTT AGCGCCGAGACG	3

Constructions were done in three steps of PCR amplification. PCR reactions were carried out following the recommendation of the manufacturer (Stratagene). The first step consisted in the construction of a *lac3* gene in which the codon 40 encoding a K was substituted by codons corresponding either to a M, a Q or a I generating the *lac3*-(K₄₀Q), *lac3*-(K₄₀M) and *lac3*-(K₄₀I) variant sequences. In a second step, plasmids containing the *lac3* variants sequences (K₄₀Q, M or I) were used as templates to substitute the codon 71 encoding a K by a codon corresponding to H, A, R or E generating combinatorial variant sequences: *lac3*-(K₄₀Q, K₇₁H), *lac3*-(K₄₀Q, K₇₁A), *lac3*-(K₄₀Q, K₇₁R), *lac3*-(K₄₀Q, K₇₁E), *lac3*-(K₄₀M, K₇₁H), *lac3*-(K₄₀M, K₇₁A), *lac3*-(K₄₀M, K₇₁R), *lac3*-(K₄₀M, K₇₁E), *lac3*-(K₄₀I, K₇₁H), *lac3*-(K₄₀I, K₇₁A), *lac3*-(K₄₀I, K₇₁R), *lac3*-(K₄₀I, K₇₁E). Sequence variations were chosen after examination of the alignment of basidiomycetes laccase sequences present in the database.

Plasmids were isolated from transformed DH5α cells and the presence of the desired mutations (as well as the absence of unwanted mutations in the rest of the sequence) was confirmed by DNA sequence analysis. The third step corresponded to the substitution of either codon 157 or codon 161 encoding a R in the previously generated variant sequences. The PCR reaction was performed with either the R157K or R161K Fwd/Rev primers couples on a mix of the 12 *lac3* double mutant plasmids (100 ng total). Subsequently, all the DH5α transformants were mixed by the addition of 2 ml of LB medium directly on the transformation plate and collected. The cell mix was used to isolate the mutagenized plasmids and 1 μg of DNA of this library was used to transform W303-1A yeast cells.¹

¹ A. Klonowska, C. Gaudin, M. Asso, A. Fournel, M. Réglie, T. Tron, *Enz. Microb. Technol.* **2005**, 36, 34.

Laccase activity assay

Standard assay

Laccase concentration was determined by the Bradford method using BSA as standard or by UV-vis spectroscopy using an $\epsilon_{600\text{ nm}} = 5 \times 10^3 \text{ M}^{-1} \cdot \text{cm}^{-1}$ for the LMCT of the T1 copper. Laccase activity was routinely assayed spectrophotometrically (Cary 50 UVvis spectrophotometer) at 30 °C using syringaldazine (4-Hydroxy-3,5-dimethoxybenzaldehyde azine (SGZ), $\epsilon_{525\text{ nm}} = 6.5 \times 10^4 \text{ M}^{-1} \cdot \text{cm}^{-1}$) as substrate as previously described.¹ One unit (U) of laccase oxidizes one μmole of substrate per minute.

Micro-titer plate assay

Each yeast colony carrying a plasmid with a *lac3* triply mutated were inoculated with a toothpick in 300 μL of SD medium supplemented with 100 μM Cu^{2+} in 96 well micro-titer plate. Cells were grown for three days at 30 °C. Then, 10 μL of cells suspensions were used to inoculate 300 μL of fresh SD medium and each culture was grown for one more day at 30°C. Cells were pelleted at 3000 rpm for 2 min and 300 μL of supernatants (SN) were collected into a clean micro-titer plate for measuring activity and protein concentration.

SNs were diluted in acetate buffer 0,1M, ABTS (2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) 5mM. SN/20 and SN/40 dilutions were directly read at 414 nm ($\epsilon_{414\text{ nm}} = 3.5 \times 10^4 \text{ M}^{-1} \cdot \text{cm}^{-1}$), with 5 seconds of shaking, at 30°C. Kinetics were followed for 1h with 30 seconds intervals with a KC4 micro-titer plate reader (BioTek, Winooski, VT, USA).

Bradford assays were performed on 150 μL of non-diluted SN. The laccase protein concentration was estimated from a standard curve obtained from BSA solution (Bovine Serum Albumin from 0 to 80 $\mu\text{g/ml}$).

Kinetic parameters

Kinetic parameters were evaluated using ABTS (Table 2) or syringaldazine (Table SI2) as substrate in B&R buffer pH 5.5 at 25°C.

Apparent K_M and k_{cat} values were obtained from the initial rate (v), enzyme concentration (E) and substrate concentration (S) according to the equation $v = k_{cat} E S / (K_M + S)$ (non-linear regression fitting using prizm program, Graph-pad, San Diego, CA). Because laccase catalysis involves two substrates and the $[\text{O}_2]$ was invariant and assumed to be saturating in this study, the measured K_M for the various substrates used should be considered apparent. Because of the assumption that 100% of the laccase participated in the catalysis as active enzyme, the measured k_{cat} should also be considered apparent.

Table SI2. Laccase variants apparent kinetic parameters on SGZ.^[a]

Enzymes	Syringaldazine		
	$k_{cat} \cdot \text{s}^{-1}$	$K_M \cdot \text{M}^{-1}$	$k_{cat}/K_M \cdot \text{s}^{-1}\text{M}^{-1}$
LAC3	71 ± 4	6.2 10 ⁻⁶ ± 1.1	1.14 10 ⁷
1-LAC3	66 ± 3	12 10 ⁻⁶ ± 1.6	5.30 10 ⁶
UNIK ₁₅₇	121 ± 4	9.6 10 ⁻⁶ ± 1.3	1.27 10 ⁷
1-UNIK ₁₅₇	65 ± 3	9.2 10 ⁻⁶ ± 1.4	7.08 10 ⁶
UNIK ₁₆₁	149 ± 5	8.7 10 ⁻⁶ ± 1.3	1.71 10 ⁷
1-UNIK ₁₆₁	62 ± 4	6.0 10 ⁻⁶ ± 1.6	1.02 10 ⁷

[a] in B&R buffer pH 5.5, 25°C; [enzyme] ca. 5 x10⁻¹⁰ M; SGZ concentration ranging from 0.8 to 80 x10⁻⁶ M.

Enzymes functionalization

The chemical modification of the enzymes was adapted from the procedure described by McFarland *et al.*² To a solution of enzymes (50 μ M final concentration) in phosphate buffer pH7.4 containing 50 mM of formate, 10 eq. of aldehyde were added as well as the Iridium catalyst (extemporaneously synthesized as described in reference 2), in a total volume of 2.5 mL. After 72h of stirring the reaction mixture was then passed over a size exclusion chromatography (Sephadex G25) in order to eliminate reagents in excess. Fractions containing the modified protein were then concentrated using a centricon (cutoff 30 kDa). During concentration the buffer was exchanged against acetate buffer 50mM pH 5,7.

Mass spectrometry

Prior digestion proteins were reduced with dithiothreitol (60-fold molar excess over disulfide bridges in 250 mM Tris-HCl, 4 mM EDTA, 6 M guanidine, pH 8.5, under nitrogen at 40°C for 20 h in the dark, followed by S alkylation with 2-iodoacetamide for 20 min at room temperature. Desalting of S-alkylated protein was done using a PD10 column. The alkylated proteins were digested with trypsin (5% w/w) for 20 h in 125 mM of NH_4HCO_3 buffer pH 8.0 at 37°C. The resulting peptides were purified on an Alliance Separation Module 2695 system (Waters) equipped with a C18 Nucleosil column (150/4). A linear gradient up to 50% solvent B was applied in 45 min, maintained at a plateau of 5 min and followed by a linear gradient up to 100% B in 5 min maintained at a plateau for 5. Solvent A: water 0.1% TFA; solvent B: acetonitrile 0.1% TFA. Elution was monitored both by UV and by fluorescence. Peaks displaying fluorescence were collected and submitted to MALDI ToF analysis. (Microflex II (2008) Bruker).

Photo reduction experiments

Light sources, filters and settings

A Dolan-Jenner MI-150 illuminator (Edmund) equipped with a 150W EKE Quartz Halogen lamp with optic fibers (\varnothing 0.8 cm) was used for oximetry, NMR and UV/VIS experiments. A power density \approx 230 mW. cm^{-2} was measured with a power meter Vector H410 connected to a Scientech head. A 450 nm OD 2 Longpass Filter (Techspec) was used to cut UV light.

Enzyme photo-reduction was carried out in Britton & Robinson buffer pH 5.0 at 25°C. Subsequent photo-catalysis experiments were performed in Britton & Robinson buffer pH 6.0 at 25°C; this pH was chosen to minimize the spontaneous opening of the epoxide ring in acidic solution while maintaining enough enzyme activity (the enzyme activity is maximum at pH 5.0).

Effect of light on laccase activity

White light irradiation induces laccase excitation with a concomitant slight reduction of the Cu(II) T1 ion observable within tens of minutes in the absence of dioxygen.³ According to the work of Henry and Peisach on the laccase from *Rhus vermicifera* (a plant laccase), this reduction is mainly attributable to the excitation of the Cu(II) T3 pair of the TNC at $\lambda = 330$ nm.⁴ In the presence of dioxygen and with the use of a 450 nm OD 2 Longpass Filter (Techspec) the enzyme photo-reduction is minimized.

Laccase activity was evaluated before and after irradiation with syringaldazine as substrate

² J. M. McFarland, M. B. Francis, *J. Am. Chem. Soc.* **2005**, *127*, 13490-13491.

³ T. Lazarides, I. V. Sazanovich, A. J. Simaan, M. C. Kafentzi, M. Delor, Y. Mekmouche, B. Faure, M. Réglier, J. A. Weinstein, A. G. Coutsolelos, T. Tron, *J. Am. Chem. Soc.* **2013**, *135*, 3095.

⁴ Y. Henry, J. Peisach, *J. Biol. Chem.* **1978**, *253*, 7751.

using standard conditions.¹ After irradiation the enzyme was found to retain 25 to 85% of its original activity depending on the length of the run.

Photoreduction kinetics

The optic fibre from the light source was adapted to the spectrophotometer's cuvette holder (90° relative to the spectrophotometer's light path) as previously described.⁵ Photo-reductions were carried out in triplicate for 1 h under irradiation with EDTA as sacrificial electron donor and several concentrations of **1**-enzyme hybrids (from 4 to 90 x 10⁻⁶ M). No significant variation of the reduction rates was observed over this range of hybrid concentrations. Reduction kinetics were fitted with a first order law (% red.= 100(1-exp(-kt))) with k being a first order rate constant. In the case of the bi-molecular system data were just interpolated

Evolution of *p*-styrene sulfonate oxidation products

Oxydation reactions were carried out in triplicate for 24 h under irradiation with 30 x 10⁻⁶ M of **1**-enzyme hybrids. In order to limit sample to sample variations, we irradiated all samples at the same time with a white LED pad (8.75 x 15 cm, ≈ 5 mW. cm⁻², Metaphase Technologies) equipped with a Wratten 2B filter λ > 450 nm.

Computational details

Homology modelling and system preparation

A three-dimensional model of LAC3 was built with homology modelling and the loops refined using Prime,⁶ choosing 3kw7.pdb as a template (75.4 % identity, 84.1% similarity, 2.9% gaps). The final structure was then prepared with the protein preparation wizard software.⁷ The protonation state of titrable residues were assigned at pH 4.0 with PROPKA⁸ and double-checked with the H++ server.⁹

Covalent Docking of complex **1** and docking of [Ru(bpy)₃]²⁺

Covalent docking was performed with Glide¹⁰ in order to graft complex **1** to K40, 71, 157 and 161, producing four **1**-UNIK variants. Although K₄₀ and K₇₁ are both bound to complex **1** in **1**-LAC3, they were treated separately for the sake of simplicity. The search space was defined as a cubic grid with a side of 30 Å centred on the grafting residue. To simulate the bi-molecular system, [Ru(bpy)₃]²⁺ was docked into the T1 site with AutoDock Vina.¹¹ In this case the search space was centered in the geometrical center of the T1 cavity, as defined with Fpocket.¹²

⁵ L. Schneider, Y. Mekmouche, P. Rousselot-Pailley, A.J. Simaan, V. Robert, M. Reglier, A. Aukauloo, T. Tron, *ChemSusChem*, **2015**, *8*, 3048-3051.

⁶ a) MP. Jacobson, RA. Friesner, Z. Xiang, B. Honig, *J Mol Biol.* **2002**, *320*, 597-608; b) MP. Jacobson, DL. Pincus, CS. Rapp, TJ. Day, B. Honig, DE. Shaw, RA. Friesner, *Proteins*, **2004**, *55*, 351-67.

⁷ GM. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, *J Comput Aided Mol Des*, **2013**, *27*, 221-34.

⁸ MH. Olsson, CR. Søndergaard, M. Rostkowski, JH. Jensen, *J Chem Theory Comput*, **2011**, *7*, 525-37.

⁹ JC. Gordon, JB. Myers, T. Folta, V. Shoja, LS. Heath, A. Onufriev, *Nucleic Acids Res*, **2005**, *33*, W368-71.

¹⁰ K. Zhu, K. W. Borrelli, J. R. Greenwood, T. Day, R. Abel, R. S. Farid, E. Harder, *J. Chem. Inf. Model*, **2014**, *54*, 1932–1940.

¹¹ O. Trott, AJ. Olson, *J Comput Chem*, **2010**, *31*, 455-61.

¹² P. Schmidtke, V. Le Guilloux, J. Maupetit, P. Tufféry, *Nucleic Acids Res*, **2010**, *38*, W582-9.

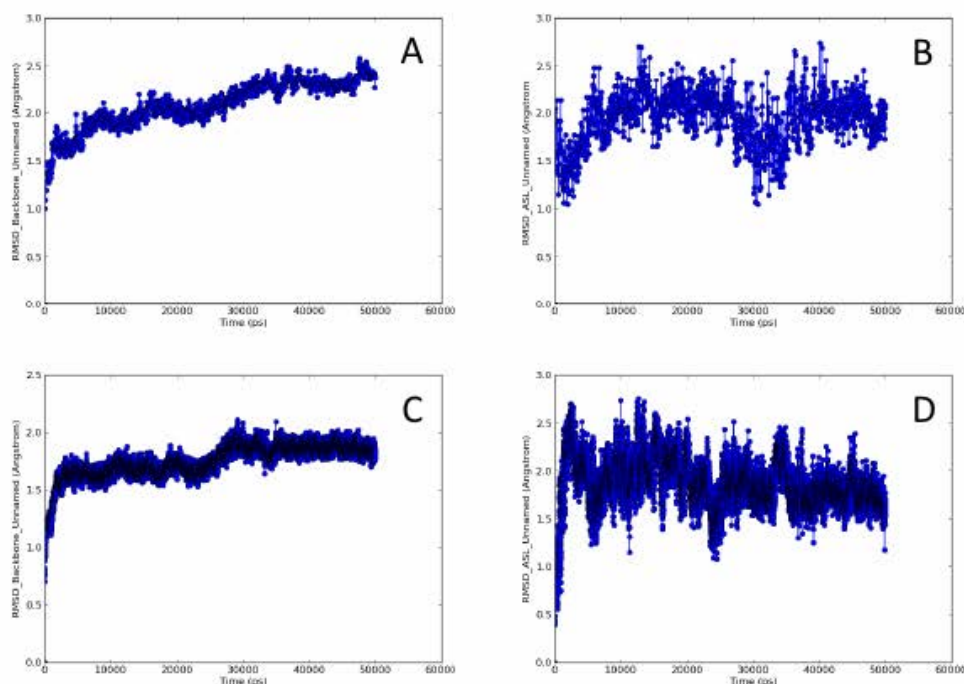


Figure S11. A: Protein backbone RMSD profile of 1-UNIK₁₆₁ 8.15 ns MD run; B: heavy atom RMSD profile of the K₁₆₁ residue modified with complex 1 in 1-UNIK₁₆₁ 8.15 ns MD run; C: Protein backbone RMSD profile of 1-UNIK₁₅₇ 50 ns MD run; D: heavy atom RMSD profile of the K₁₅₇ residue modified with complex 1 in 1-UNIK₁₅₇ 50 ns MD run.

Molecular dynamics and metadynamics

After covalent docking, 1-UNIK variants were solvated with a 10 Å buffer of water molecules in an orthorhombic box, neutralized and 0.15 M NaCl was added. After equilibration (default settings), a 50 ns long NPT production at 300 K was performed with Desmond,¹³ using the OPLS-2005 force-field¹⁴ and the SPC explicit water model.¹⁵ The temperature was regulated with the Nosé-Hoover chain thermostat¹⁶ with a relaxation time of 1.0 ps, and the pressure was controlled with the Martyna-Tobias-Klein barostat¹⁷ with isotropic coupling and a relaxation time of 2.0 ps. The RESPA integrator¹⁸ was employed with bonded, near, and far time steps of 2.0, 2.0, and 6.0 fs, respectively. A 9 Å cut-off was used for non bonded interactions together with the smooth particle mesh Ewald method.¹⁹

¹³ K. J. Bowers, E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan, D. E. Shaw, *SC Conference*, **2006**, sc06.supercomputing.org/schedule/pdf/pap259.pdf.

¹⁴ G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, *J. Phys. Chem. B*, **2001**, *105*, 6474-6487.

¹⁵ K. Toukan, A. Rahman, *Phys Rev B Condens Matter*, **1985**, *31*, 2643-2648.

¹⁶ S. Nosé, *J. Chem. Phys.*, **1984**, *81*, 511-519.

¹⁷ G. J. Martyna, D. J. Tobias, M. L. Klein, *J. Chem. Phys.*, **1994**, *101*, 4177-4189.

¹⁸ M. Tuckerman, B. J. Berne, G. J. Martyna, *J. Chem. Phys.*, **1992**, *97*, 1990-2001.

¹⁹ U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577-8593.

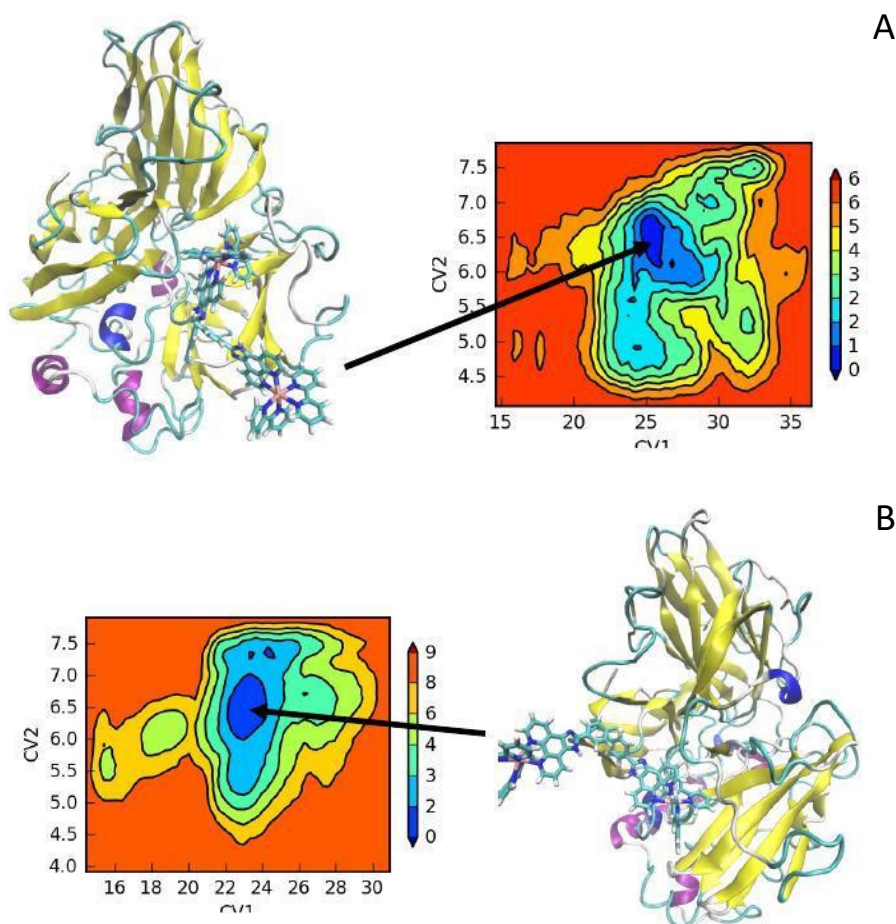


Figure S12. Metadynamics results for A: 1-UNIK₄₀ and B: 1-UNIK₇₁. CV1 is the T2-Ru distance, CV2 is the CA-NZ distance.

In order to assess if complex **1** tends to stay on the surface, highly exposed to water, or to hide into the protein interior (as covalent docking suggested) of LAC3, 1-UNIK₄₀ and 1-UNIK₇₁ were simulated with metadynamics for 20 ns.²⁰ It should be noted that K₄₀ and K₇₁ were treated separately (i.e. replacing the second K residue by a serine (S)) to keep the number of collective variables (CVs) tractable. Two distances were chosen as CVs: (i) between Ru and T2 copper and (ii) between CA and NZ atoms belonging to the lysine. The free energy surface (FES) was constructed depositing a gaussian every 0.09 ps. The width of each gaussian was 0.05 Å and the height 0.03 kcal/mol. Two distinct simulations were carried out for each protein variant.

Electronic coupling calculations

Electronic couplings have been estimated using the VMD Pathways plugin.²¹ In the Pathways model, TDA is quantified following continuous paths between donor (the Ru atom) and acceptor (each Cu atom), where the electron tunnels through space, a hydrogen bond or a covalent bond. The number and nature of tunnelling events determines the electronic

²⁰ A. Laio, M. Parrinello, *Proc Nat. Acad. Sci. USA*, **2002**, 99, 12562–12566.

²¹ IA. Balabin, X Hu, DN. Beratan, *J Comput Chem*, **2012**, 33, 906-910.

coupling value.

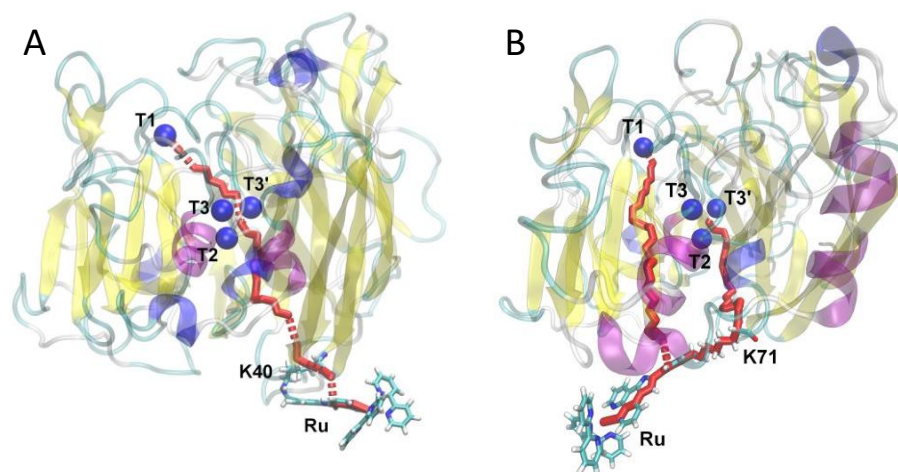


Figure S3. Position of Complex 1 in A: 1-UNIK₄₀ and B: 1-UNIK₇₁ as a simplified models of 1-LAC3 obtained from covalent docking plus metadynamics simulations. Electron transfer pathways are depicted in red.

IV. DISCUSSION

1. Results, summary and discussion

Since this thesis is presented as a compendium of articles, the work performed by myself is not clearly separated. Therefore, in order to showcase my personal contribution to these manuscripts and avoid getting credit for the work performed by others, only results obtained by myself will be presented in the following section. In the case where the isolation of my contribution was not feasible, due to collaborations with other members of the lab, they will be properly mentioned and acknowledged.

Under the framework of the INDOX (INDustrial OXidoreductases <http://www.indoxproject.eu/>) European project, where the final aim was to provide relevant industrial cases to demonstrate the efficiency of biocatalyst on targeted reactions, the stories here presented include functionalized polymers for pulp and paper industries (lignin degradation) and precursors for specialty polymers (HMF products). The expertise that the group acquired opened the possibility of a collaboration to further work in laccases which could be included in this thesis since laccases have industrial applications for the pulp and paper industries too. These works include:

1. Selective oxidation of secondary alcohols by means of flavoproteins

- Rationalization of F397 (*Computer simulations were performed by Pep Amengual Rigo, a master's student at the time, and the work was **supervised** by me*)
- Combinatorial saturation mutagenesis for AAO residues 500 and 501
- Rationalization of the mutations introduced in FX9 variant of AAO
- Electrostatics evaluation per residue to increase the FAD redox potential

2. Improvement of full HMF oxidation (*UPO studies performed by Marina Cañellas*)

- Study of AAO in HMF oxidation steps
- HMFO as an alternative flavoprotein

3. Laccases (*in collaboration with Emanuele Monza*)

- Design of improved UNIK variants
- Study of LAC3 grafting in an electrode

1.1. Secondary alcohol oxidation by means of flavoproteins

The main goal of this project was to engineer the AAO protein in order to oxidize stereoselectively the (S)-1-(p-methoxyphenyl) ethanol (SMP) over (R)-1-(p-methoxyphenyl) ethanol (RMP) from the racemic mixture. This is basically the secondary alcohol of the native substrate of this enzyme, p-anisyl (ANI). The activity of the enzyme was increased for the F501A variant, removing the steric hindrance between the phenylalanine with the methyl group of the secondary alcohol (Hernández-Ortega et al. 2012).

Therefore, the strategy followed in order to further increase the selective oxidation towards SMP consisted of simulating the ligand diffusion of both enantiomers of the racemic mixture (SMP and RMP) in the F501A variant of the AAO, by means of PELE. The information extracted from these simulations included the position adopted for each ligand in the AAO's active site and a contact map. This was built by defining an arbitrary threshold of 2.6 Å, based on the approximate sum of a hydrogen and heavy atom radius: when the distance between any atom of the substrate and an atom of any residue was under this threshold along PELE stochastic trajectories, a contact was added to the counter, providing at the end a frequency of interaction between residues and each ligand. The following figures show the number of contacts per residue for SMP (Figure 16) and RMP (Figure 17).

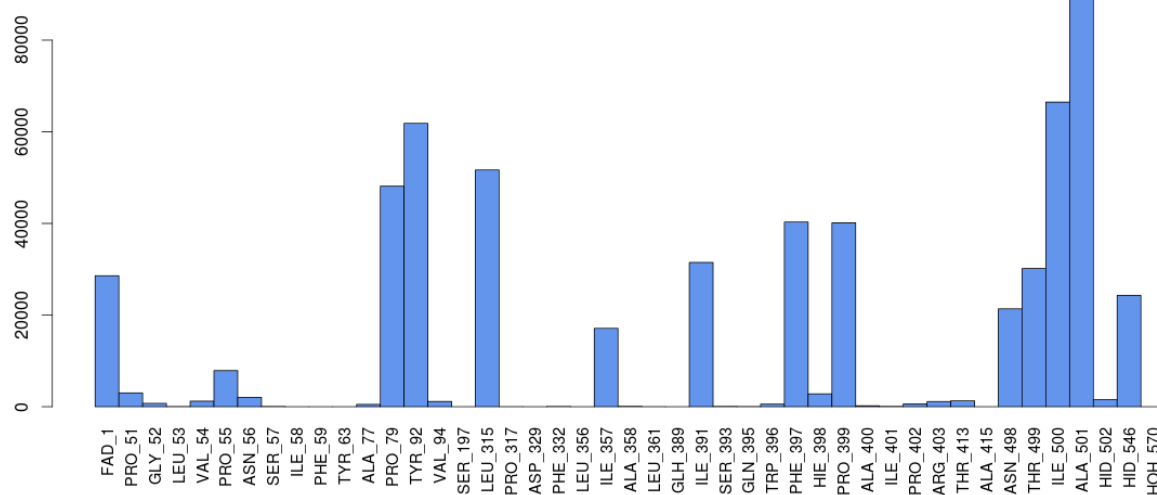


Figure 16. Contact map between **SMP** ligand and F501A AAO variant. Contacts are considered when distance between two atoms is below 2.6 Å.

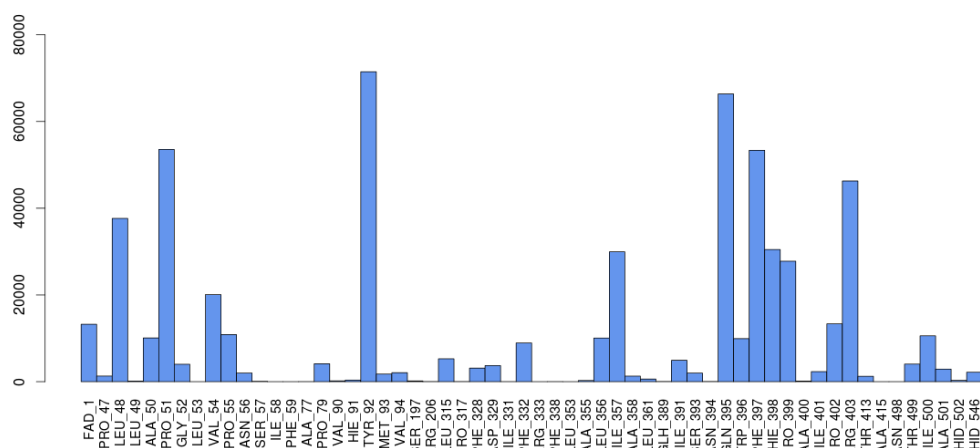


Figure 17. Contact map between *RMP* ligand and F501A AAO variant. Contacts are considered when distance between two atoms is below 2.6 Å.

Rational design of the active site was assayed: providing that even though an alanine is one of the smallest amino acids, it was still the residue with the largest number of contacts. Therefore, in an attempt to reorder the backbone of the region, the variant including F501P and I357A was simulated, showing promising results. Figure 18 shows how PELE produced many more conformations with better catalytic distances for the F501P-I357A variant. However, the introduction of a proline in the active site is always risky, but this was considered a high-risk high-reward mutation.

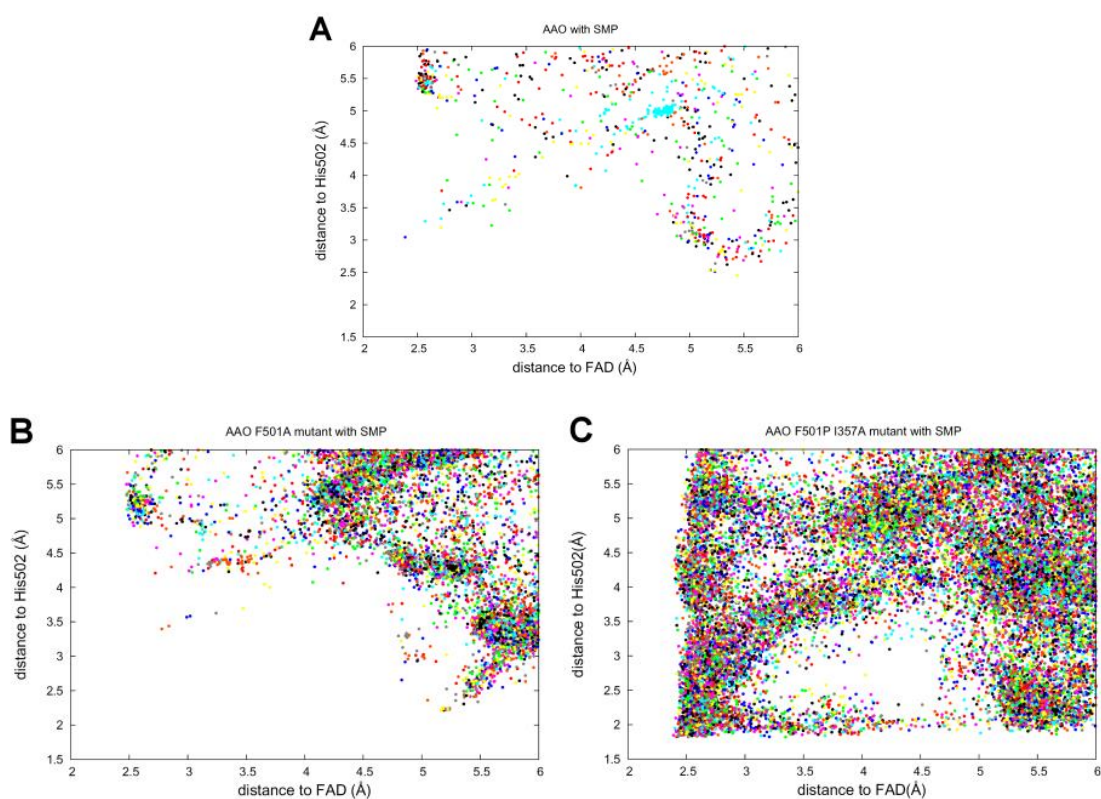


Figure 18. Plots represent PELE simulations relating catalytic distances (X and Y) of SMP for different AAO variants, including A) wild-type, B) F501 variant and C) F501P-I357A variant.

This was suggested to the experimental lab, but the FAD could not be incorporated, and therefore no activity was shown. However, the contact map revealed that i) the highest number of contacts corresponded to positions to residues I500 and A501, and ii) the residues involved in the formation of a hydrophobic bottleneck (Tyr92, Phe501 and Phe397) presented a relatively elevated number of contacts even after the substitution of a bulky phenylalanine by an alanine in the F501A variant. The collaborators from CSIC Madrid, the experimental group led by Professor Angel T. Martínez, designed saturated mutagenesis experiments for both cases.

1.1.1. F397 variants for ligand diffusion

The experimental variants F397Y, F397W, F397A and F397L were characterized and compared to the native AAO by computational means, to reveal the role of F397 in the catalytic process. MD simulations showed that loops Gln395-Thr406 and Ser89-Met95 (containing gate residues Phe397 and Tyr92) presented low RMSD values during the simulation, rejecting the hypothesis of a large conformational change. However, it could be observed how F397Y and F397W variants offered the possibility of a hydrogen bond formation with Tyr92, displaying shorter distances between the gate residues compared to the aliphatic variants (Figure 19).

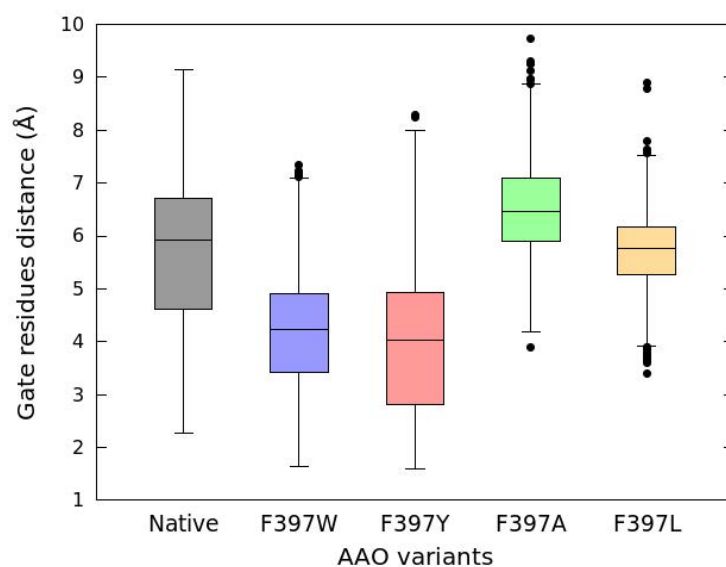


Figure 19. Boxplot of the distance between gate residues of the different systems.

Moreover, PELE simulations reproduced the entrance of the substrates (ANI and O₂) to the active site and the product release for the oxidized *p*-anisic acid and *p*-anisaldehyde.

Energy profiles revealed a minimum at the gate only for acidic products. Moreover, the average number of total PELE steps needed to release the products is larger for the acidic products, especially for the aromatic variants. On the other hand, the diffusion study of the reactants shows similar results for every variant except for F397W, which present a higher population at states with shorter distances between O₂ and FAD (Figures 21 and 22). For the alcohol substrate, less efficient catalytic poses are found, due to the presence of an alternative minimum caused by hydrogen bond interactions between ANI's hydroxyl and Trp397 (Figure 20).

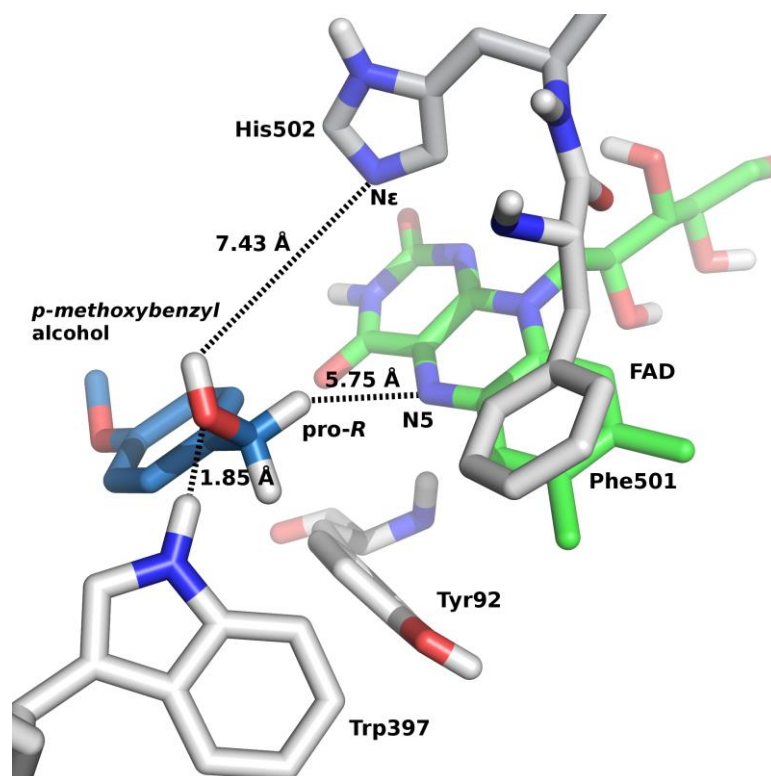


Figure 20. New ANI conformation only present in the F397W variant, blocking the ligand diffusion.

Even though previous computational studies already hinted that Phe397 could oscillate with the substrate as a gating mechanism, these simulations (together with the experimental results) confirmed the important role of Phe397. Besides its participation in the ligand diffusion, results prove that it also favours a correct positioning of the alcohol for the oxidation, evidenced by F397W results. This position has also an influence in the oxidative half-reaction, improving the interaction of the enzyme with O₂.

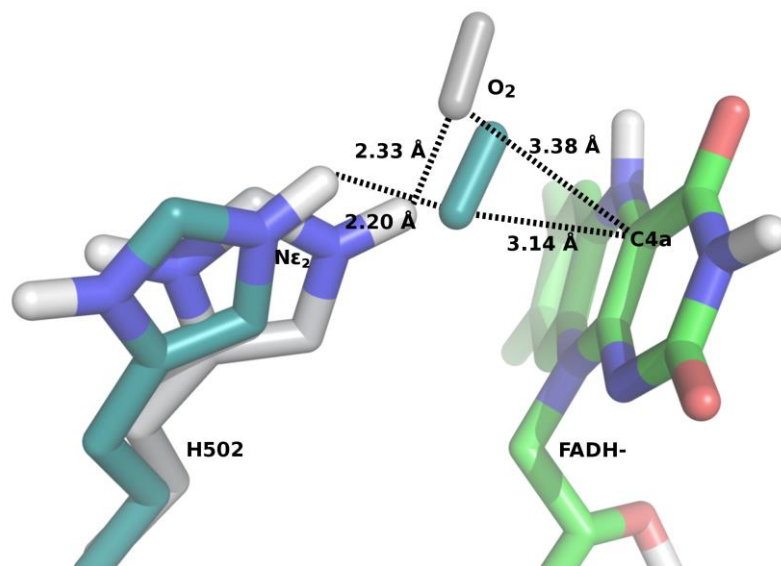


Figure 21. Difference in oxygen catalytic positioning for the native AAO (grey) and F397W variant (green). Shorter distances for F397W variant indicate and improvement in oxygen positioning.

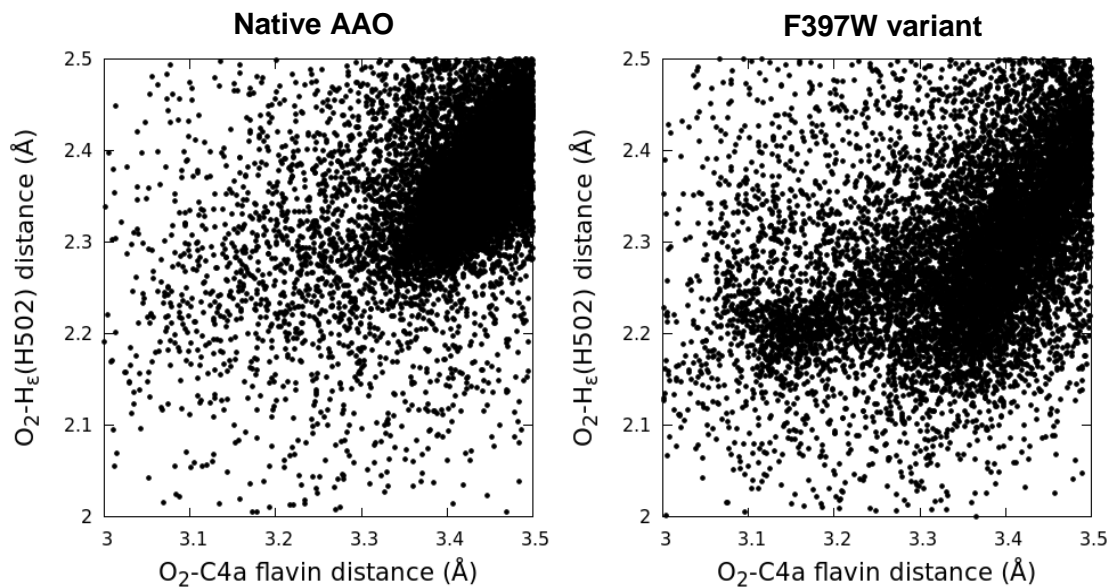


Figure 22. Catalytic distance distribution between O₂, FAD and His502 for the native AAO (left) and the F397W variant (right).

1.1.2. Combinatorial mutagenesis for residues 500 and 501 in AAO

After the experimental obtention of successful variants to increase oxidation of secondary alcohols (up to 100-fold increment in catalytic efficiencies) in the positions with the largest number of contacts for SMP, the effect of each variant was modelled and rationalized computationally.

The effect of these mutations was evaluated with PELE, for both SMP and RMP substrates. Initially, the assumption was that the main effect could be found in the half-reductive reaction since it is the limiting step. The representation of both catalytic distances in Figure 23, being x) the distance between the alcohol hydrogen of the ligand and the histidine's nitrogen and y) the distance between the hydride and the FAD N5 (explained in detail in section 2.2 *Flavoproteins*) provides a simplified picture of the ligand orientation in the active site. Clearly, the difference between reactive SMP (left) and non-reactive RMP (right) can be appreciated, since the latter cannot reach catalytic distances lower than 3Å for both distances at the same time (notice that the reaction follows a concerted mechanism), with good interaction energies.

Differences in activity for SMP among the different variants are not so straightforward to identify. Subtle changes in catalytic distances can reduce k_1 (following the Michaelis-Menten notation in equation 1), but also interaction energy variations, the population of catalytically relevant states or conformation accessibility can affect the kinetic constants. Clustering the catalytic efficiencies into three groups (low efficiencies for AAO and F501, intermediate for I500A and high efficiencies for I500M and I500M/F501W), it can be qualitatively correlated with Figure 23. Higher efficiencies present shorter distances compared to the group with low efficiencies. Interestingly, I500A show very different behaviour. The plot does not show a discrete set of distances between the FAD and the substrate when the latter is in contact with the catalytic histidine, His502. Compared to F501A, both present similar minimum catalytic distances, but I500A would access to these states in an easier manner, due to the capability of freely move in the active site.

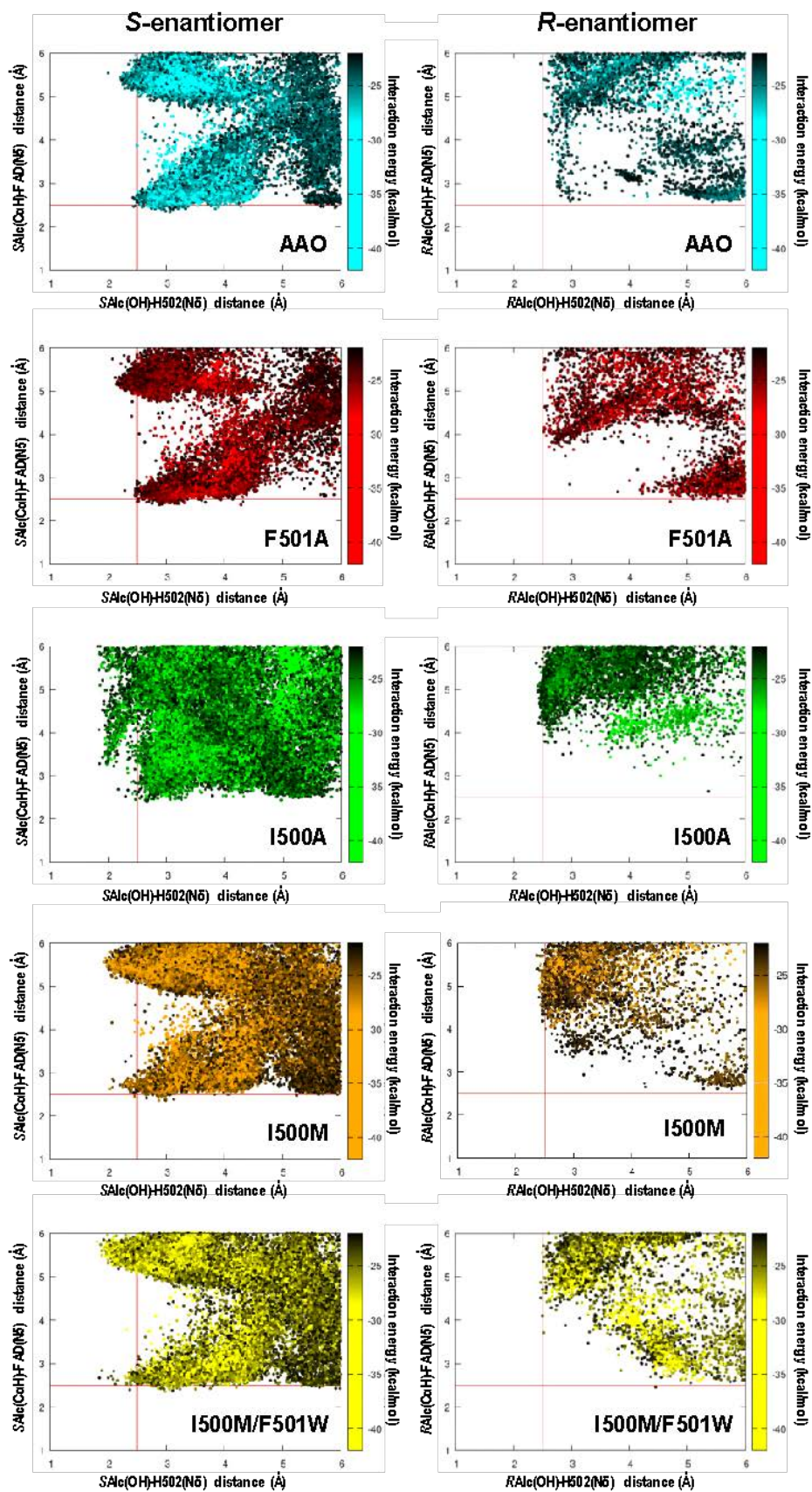


Figure 23: Plots represent PELE simulations relating catalytic distances (X and Y) and interaction energies (color scheme, right Y axis) for different AAO variants.

1.1.3. Rationalization of the mutations introduced in FX9 variant of AAO

Another collaborator from Madrid, Miguel Alcalde's lab, reached a similar conclusion starting from the FX9 variant of AAO (which includes mutations H91N and L170M). After combinatorial saturated mutagenesis of positions Ile500 and Phe501, different variants were obtained with improved activity towards secondary benzyl-alcohol (coinciding with those from the previous section) being I500M-F501W the best mutant, with an improvement of 160-fold compared to the native FX9. Here, it is important to mention the effect of the mutation at position 91, which is found in the active site, close to the FAD N5 atom. By itself, this mutation seems to increase the activity of AAO, and it was concluded experimentally that its effect was related to the enzyme expression. However, it may also have an effect on the FAD positioning and, importantly, on its reactivity through the N5 hybridization (planar or tetrahedral) that might directly affect its redox potential.

Subsequent rounds of random mutagenesis further improved this enzyme for secondary benzyl-alcohols. In particular, the variant named 7A9 (that included mutations H91N, L170M, A77V, R80C, V340A, I500M and F501W) was computationally studied. Experimental results were successfully reproduced with simulations, where PELE was able to find significant differences in the orientation of the substrate (in this case, SMP), showing that mutations allowed the ligand to position the reactive hydrogens towards the His502 and FAD respectively, with significantly shorter distances than the native protein (Figure 24).

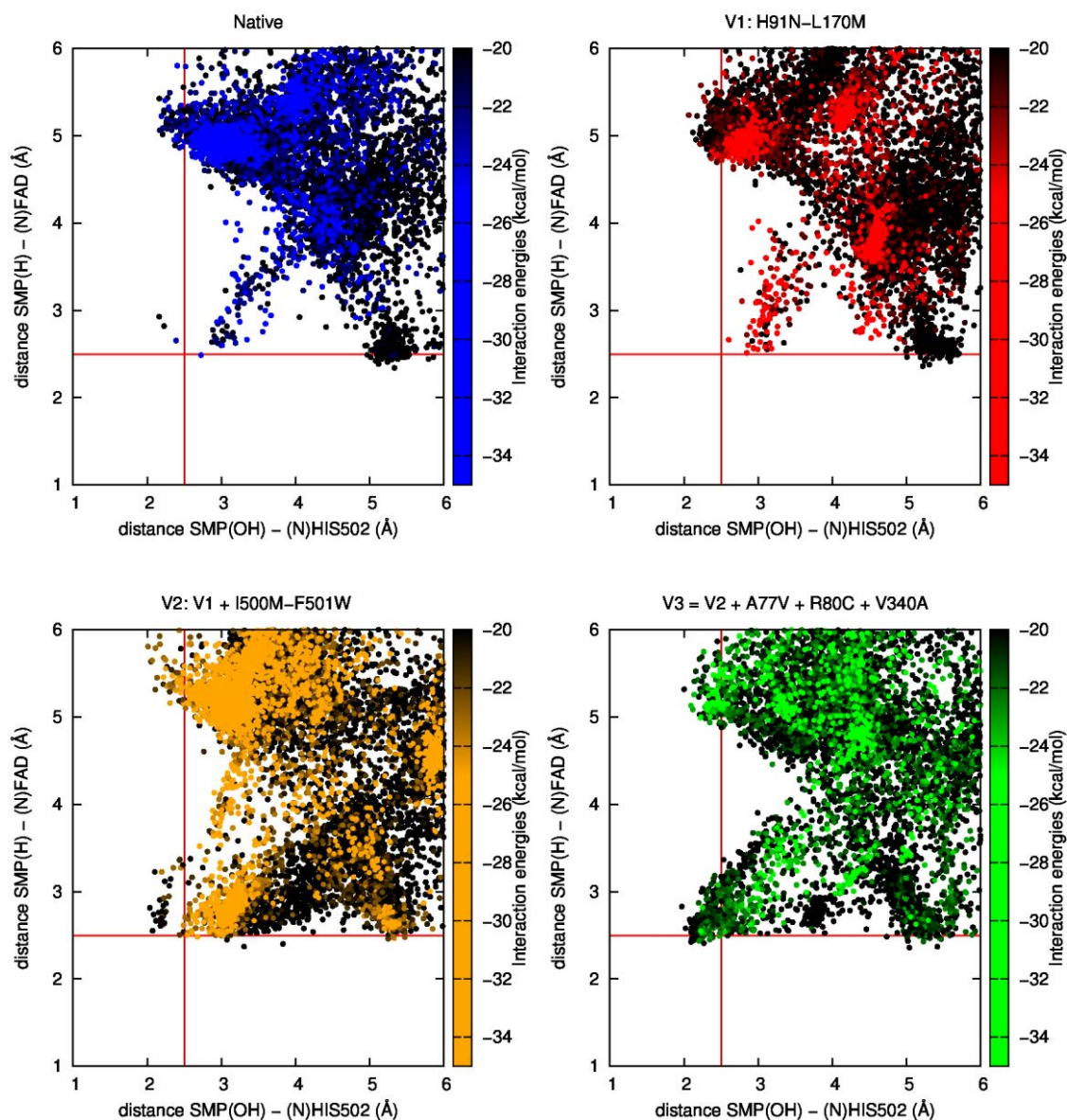


Figure 24. Plots represent PELE simulations relating catalytic distances (X and Y) and interaction energies (color scheme, right Y axis) for different AAO variants.

The large number of mutations introduced in this variant difficult the rationalization of the effect of each mutation. For this reason, the resulting variants after each round were simulated. It is important to mention that some of these mutations are not located near the active site cavity, making its contribution to the catalysis improvement not obvious. Their effect is assigned to enhancements in the expression of the protein, without a direct relation with the enzymatic cycle. This would be the case of L170M and V340A long-distance mutations. But the contribution of the rest of the mutations is illustrated in the following figures:

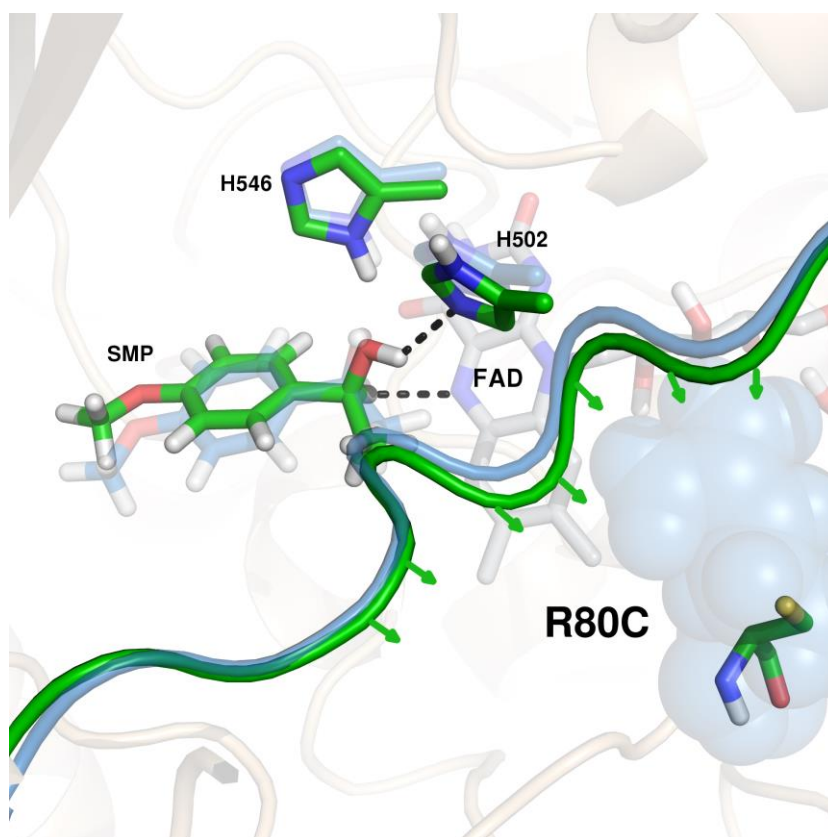


Figure 25. Structural differences of SMP catalytic position for AAO (transparent) and 7A9. The displacement of the backbone due to R80C mutation is shown.

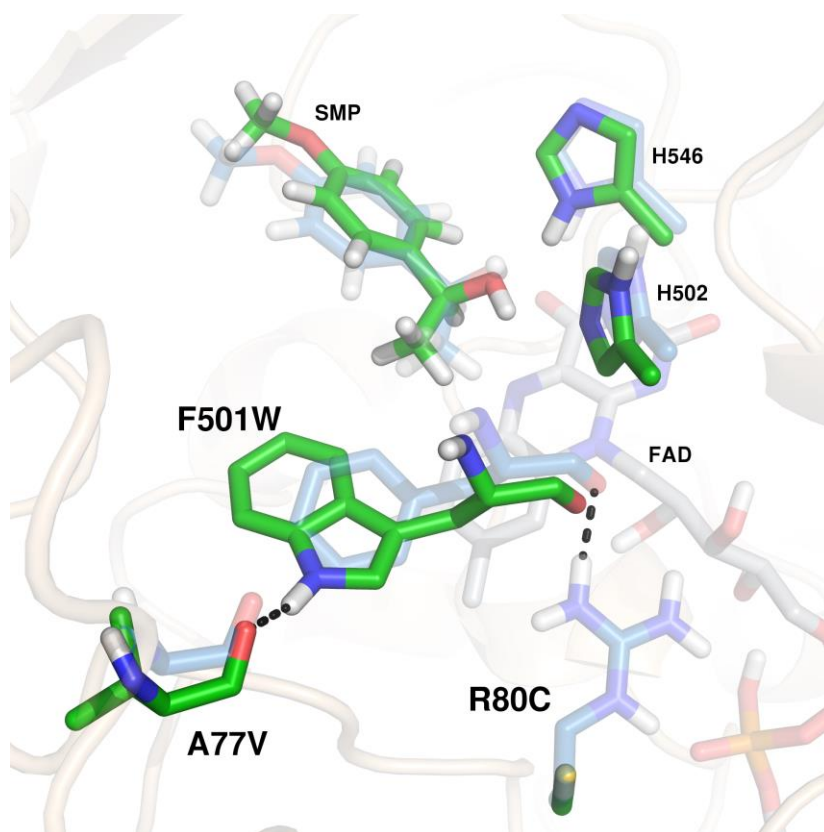


Figure 26. Structural differences of SMP catalytic position for AAO (transparent) and 7A9. Differences in interactions with residue 501 are shown.

As illustrated in Figure 25, the substitution of Arg80 creates an empty space displacing the backbone where the important residues 501 and 502 (the catalytic histidine, that acts as a base) are present (see chapter 2.2 *Flavoproteins* for more mechanistic details). Moreover, mutations F501W and A77V introduce a chain reaction: i) R80C mutation removes the hydrogen bond with the backbone of residue 501, ii) A77V mutation slightly displaces this residue and iii) F501W forms a hydrogen bond with Val77 oxygen backbone. These effects can be seen in Figures 26. This subtle repositioning of the backbone due to R80C mutation is crucial for the activity improvement, affecting both residues 501 and 502, and shortening the catalytic distances, improving the limiting step of the chemical transformation.

This is a good example of how residues present in the second shell of the active site can be very important for protein engineering: even if they do not interact directly with the substrate or with residues involved in catalysis, they can still originate changes to enhance the protein activity.

1.1.4. Tuning FAD potential by the evaluation of residue electrostatics

In parallel, the development of an algorithm to evaluate the electrostatic contribution of each residue to the reaction was assayed. The goal was to identify long-range hotspots to mutate and increase the reaction rate without modifying the architecture of the active site. The algorithm consists in calculating the energy barrier of a reaction, which implies the laborious task of optimizing the transition state (TS). Then, after removing the side chain charges of an amino acid (that is, to change all the side chain partial charges to zero), single-point energy calculations were performed for the reactants and the TS. Therefore, a new energy barrier could be calculated without considering the effect of the partial charges of that amino acid, and the energy barrier difference could be representative of the electrostatics effect of that residue. This algorithm was applied to the AAO-ANI protein-ligand system. The following image shows in blue those amino acids which resulted in a lower energy barrier after removing their partial charges, while in red residues that increase the energy barrier are shown.

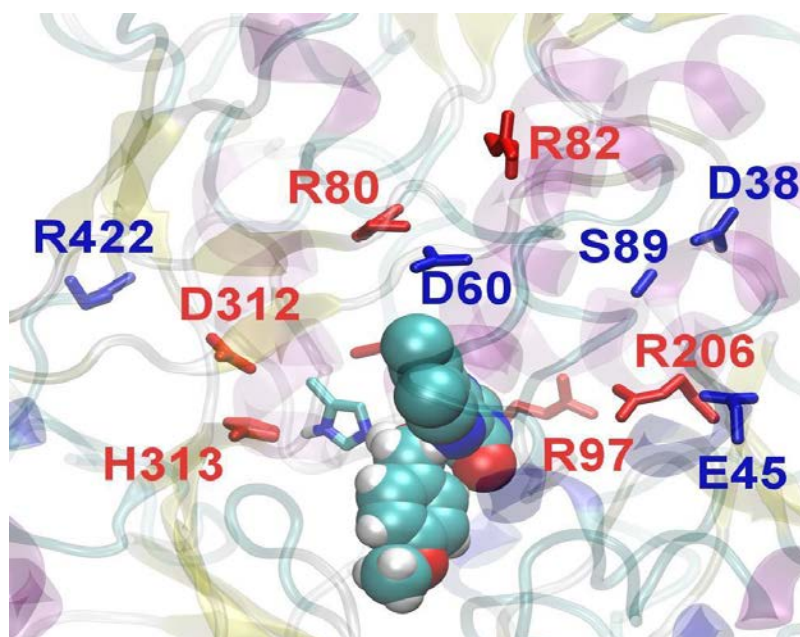


Figure 27. Representation of residues decreasing (in blue) and increasing (in red) the energy barrier when no electrostatic interactions are considered.

Some variations were derived, such as adding a positive charge or a negative charge to check for their influence over the energy barrier. After the addition of a punctual charge (positive and negative), two differentiated regions could be observed, one stabilizing the energy barrier through positive charges and the other with negative charges (Figure 28).

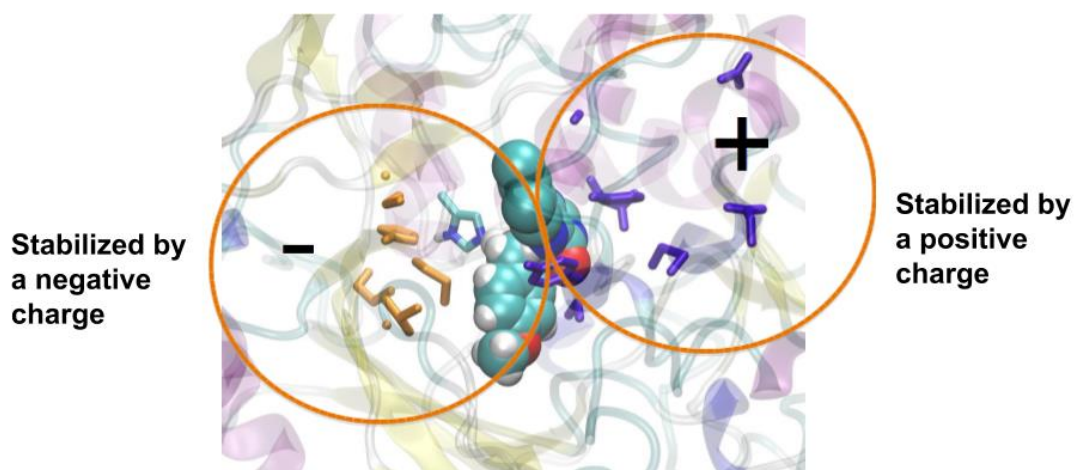


Figure 28. Two different regions can be appreciated, one stabilized by positively charged and the other by negatively charged residues.

After locating the positions where electrostatic effects could be important, five different variants were chosen, considering the charge that would benefit the most to the energy barrier and, importantly, a sequence criterion: which amino acids are present in each position in other family members. The next table summarized the selected variants.

Table 3. Resume of the suggested variants, including the score for the energy barrier change and the residues present in the same position in other proteins.

Charge	Barrier	Residue	Mutability	Proposed variant	Residue variability
0	-2.44	GLU45	8	E45Q	9xH, 8xQ , 5xR, 4xE, 3xI, 3xS, 2xM, 2xV, 1xA, 1xG, 1xK, 1xL, 1xN, 1xT
0	-1.91	ASP38	2	D38N	19xD, 12xN , 4xH, 2xA, 2xG, 1xE, 1xP, 1xY
1	-3.9	GLU45	8	E45R	9xH, 8xQ, 5xR , 4xE, 3xI, 3xS, 2xM, 2xV, 1xA, 1xG, 1xK, 1xL, 1xN, 1xT
-1	-2.94	LEU498	9	L498D	10xG, 7xS, 6xD , 4xA, 4xT, 2xH, 2xI, 2xL, 2xN, 2xQ, 1xC, 1xE, 1xF, 1xR
-1	-3.36	LEU315	9	L315D	11xL, 5xD , 5xT, 4xA, 4xF, 4xM, 4xQ, 4xV, 3xI, 2xR, 1xC, 1xG, 1xN, 1xW

After testing the suggested variants in the lab, results from Table 4 were obtained. Unfortunately, even though a higher k_{cat} was obtained for E45Q, none of these variants presented higher efficiencies for the native substrate than the wild type, and these results were not further analysed. But the potential to identify long-range hotspots influenced in the decision of including it to this thesis. Also, there is still room for improvement: a punctual charge was added to only an atom of the side chain, which could have increased the overestimation problem. The distribution of the charge through the side chain is something to be considered in the future.

Table 4. Catalytic constants obtained for the suggested variants.

<i>p</i>-methoxybenzyl alcohol activity			
	k_{cat} s^{-1}	K_M μM	k_{cat}/K_M $(mM^{-1}s^{-1})$
WT	115.5	24.7	4700
E45Q	142.9 ± 2.2	33.4 ± 1.9	4284.1 ± 185.0
E45R	98.0 ± 2.1	22.3 ± 2.0	4390.3 ± 323.4
L315D	94.9 ± 1.9	341.9 ± 20.3	277.7 ± 12.0
N498D	104.5 ± 2.3	27.0 ± 2.8	3866.6 ± 334.8

1.2. Improvement of full HMF oxidation

The objective of this project was to investigate a possible path for the transformation of HMF to 2,5-furandicarboxylic acid (FDCA). A pathway was suggested that involved AAO and one unspecific peroxygenase (UPO). Basically, due to the impossibility of catalysing the last step of the reaction for AAO, a UPO was coupled in a way that AAO oxidized the first two steps (from HMF to FFCA) and provided hydrogen peroxide to the UPO, for the oxidation of FFCA to FDCA. The whole reaction chain is summarized in Figure 29 This work, pending of publication, was done in collaboration with M. Cañellas, who was

responsible of the processes involving the UPO. Therefore, in this section, only AAO related simulations will be included.

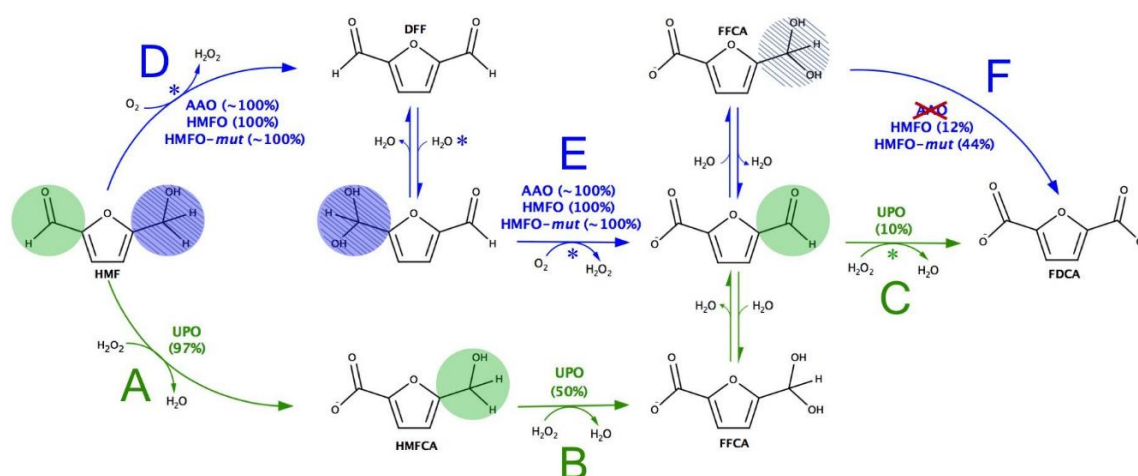


Figure 29. Complete reaction mechanism of HMF via flavoproteins, either AAO or HMFO (in blue) and UPO (in green).

First, the target was set towards the understanding of the reactivity of each substrate. More specifically, rational explanation on why FFCA was not oxidized by AAO was the main concern. Hence, a detailed evaluation was performed through substrate migration towards the active site, by means of PELE simulations.

The migration of all the substrates involved in the oxidation of AAO to FDCA was studied using PELE ligand migrations. This means that AAO, DFF and FFCA were studied. It has to be noticed that DFF and FFCA were modelled in their gem-diol form since it is the reactive one. As it can be seen in Figure 30 (A, B), similar energy profiles are obtained for the first two steps of the complete reaction. However, FFCA (Figure 30, C) is not able to reach distances lower than 4\AA , making the hydrogen abstraction energetically not possible under these conditions. Since there are no significant differences in the gem-diol group of the different substrates, the carboxylic group (or in particular its negative charge) could be the cause of the non-reactivity. Interestingly, ligand diffusion of 2,5-hydroxymethylfurancarboxylic acid (HMFCA), the product of the HMF oxidation via UPO which also has a negative charge, produces a similar energy profile than FFCA (Figure 30, D). Ligand diffusion with PELE does not produce conformations of HMFCA in which the distance between the ligand and the FAD are lower than 4\AA .

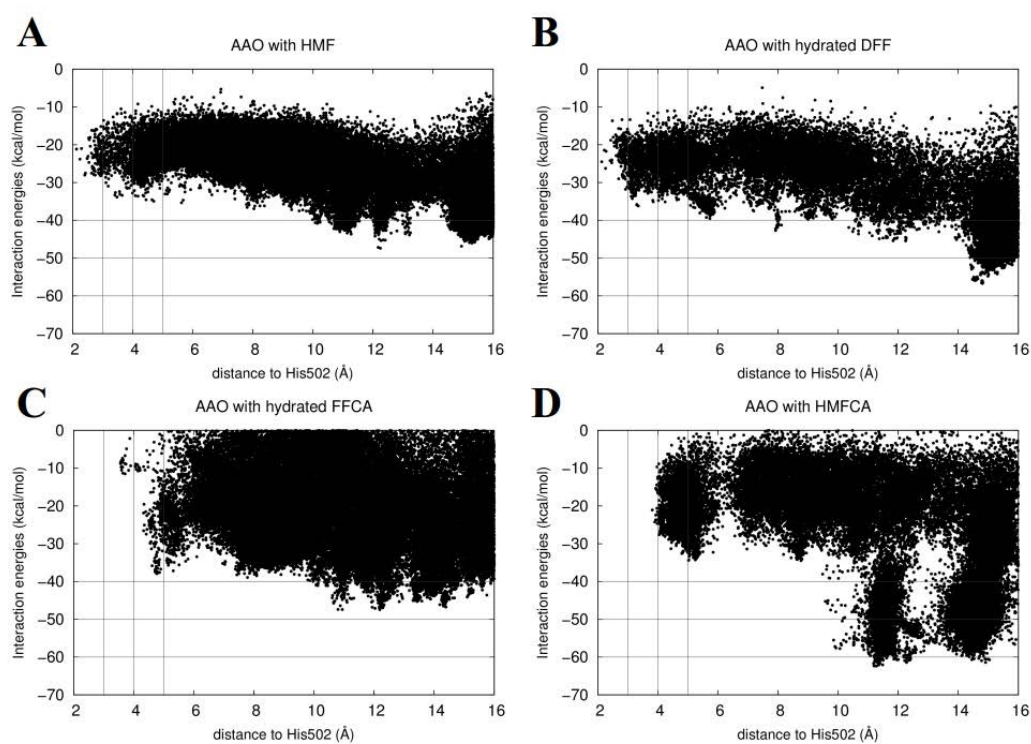


Figure 30. PELE interaction energies vs distance between N atom in the His502 and the alcohol hydrogen atom from the ligand in AAO. The plots correspond to the migration of different ligands: A - HMF, B - DFF, C - FFCA, and D - HMFA.

Visual inspection of the energy minimum shows strong interactions of the carboxylic group with the protein, as illustrated in Figure 31. In particular, residues Tyr92, His398, Gln395 and Arg403 are involved in FFCA stalling at 4Å between the hydrogen that has to be transferred, and the FAD.

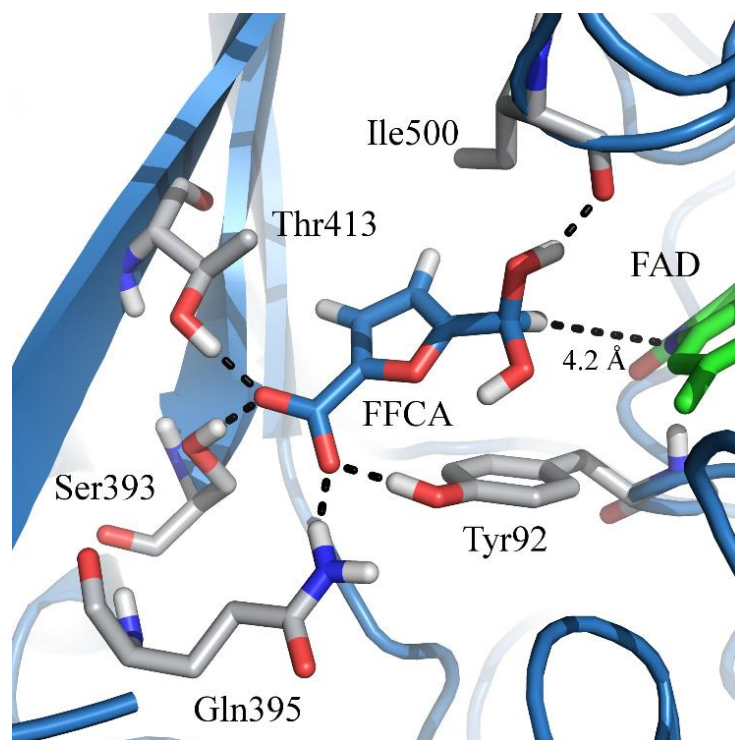


Figure 31. Representation of FFCA position in the energy minimum, where multiple interactions of the negative charge can be appreciated.

Diffusion impediments, together with low presence of the hydrated form in the aldehyde-gem-diol equilibrium and a higher substrate redox potential when it is negatively charged, could contribute to the non-reactivity of FFCA in AAO. This problem could be solved by either designing AAO variants to potentially perform all the HMF oxidation steps or by using a different protein, capable of producing FDCA from HMF. Since no AAO variants were obtained through rational design to potentially oxidize the last step of the reaction, hydroxymethylfurfural oxidase (HMFO), which was the first enzyme ever reported to catalyze the entire reaction chain from HMF to FDCA (Martin et al. 2018; Dijkman et al. 2015; Pickl et al. 2018), was suggested as an alternative enzyme. An HMFO variant, that includes the V367R/W466F double mutation, increase the activity of this last step. Ligand diffusion simulations with PELE were performed also on this system, in order to compare them, rationalize the effect of the double mutation and try to mimic the effect of these mutations on AAO.

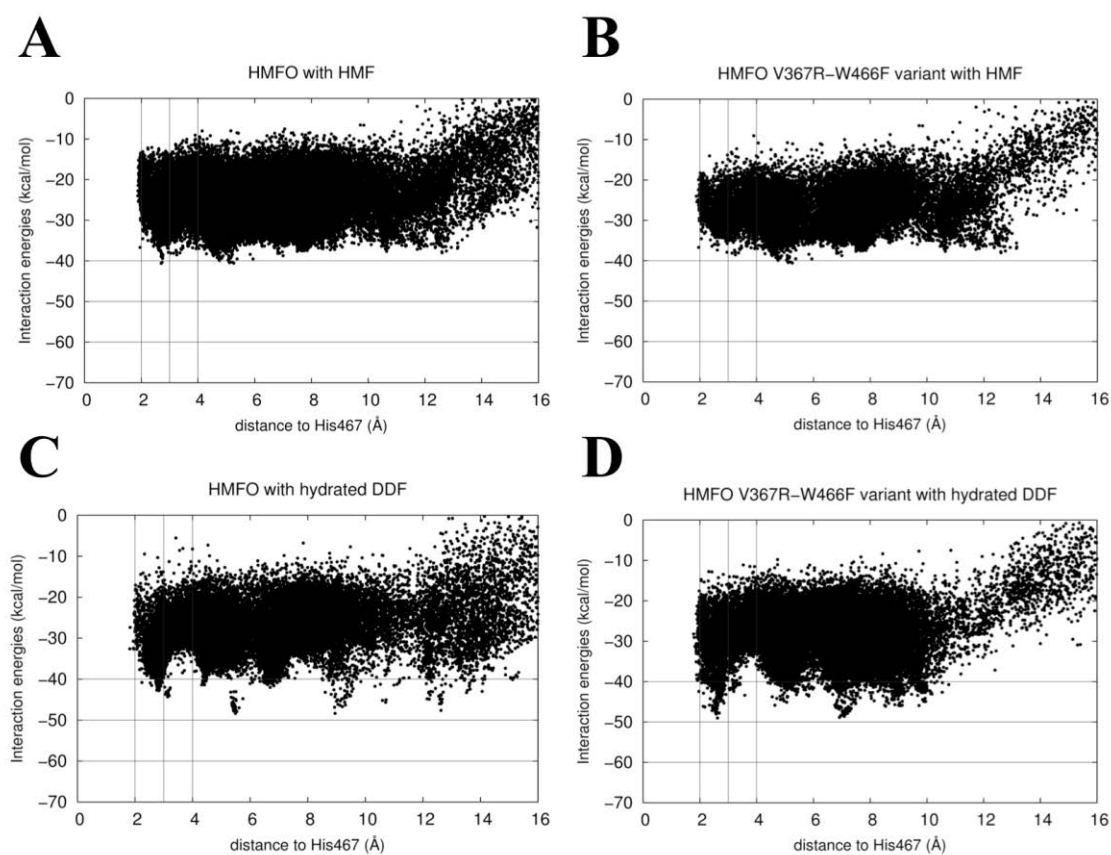


Figure 32. PELE interaction energies vs distance between N atom in His467 and the alcohol hydrogen atom from the ligand in HMFO (left) and HMFO V367R/W466F variant. The plots correspond to the migration of different ligands: A and B – HMF; C and D – hydrated DDF.

The diffusion of the substrates for the first two steps of the HMF chain reaction is shown in Figure 32. As in AAO, both HMF and DFF substrates can reach the active site with optimal catalytic distances for the proton extraction. Moreover, the introduction of the V367R/W466F double mutation barely modifies the interaction energy profiles for these substrates and thus, their oxidation is not affected. However, in the next picture, a representation of the negatively charged FFCA interaction maps is presented for HMFO and the doubly mutated variant.

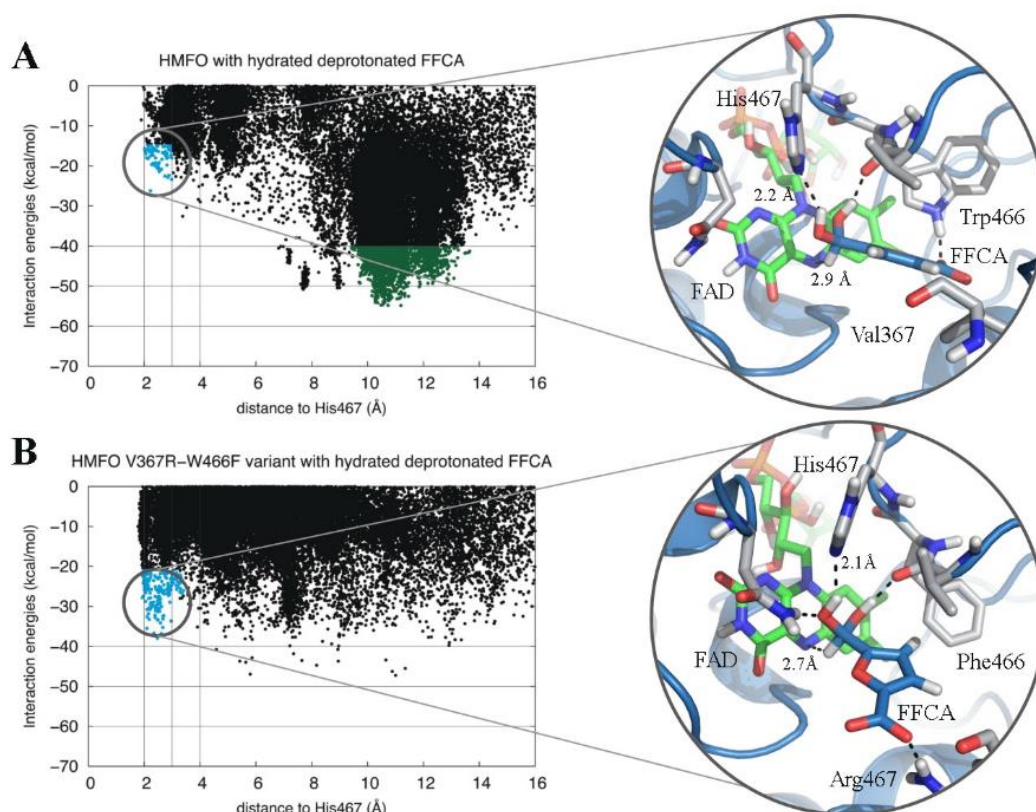


Figure 33. Interaction energy vs. distance between N atom in His467 and the alcohol hydrogen atom from FFCA in the native HMFO (A) and HMFO doubly mutated variant (B). A representative structure from the catalytic minimum is shown for each case.

Unlike in AAO plots, where FFCA was not able to get distances short enough for the proton transfer to be energetically possible, here values between 2 Å and 3 Å were obtained. The double mutation seems to decrease the interaction energy in the binding site, favouring this position. As shown in the representative structure of each minimum, this energetically preferred modification is due to: i) W466F reduces the size of the amino acid, removing the -NH group and thus, the polar interaction with the carboxylic group of the substrate, and ii) V367R introduces a positively charged residue to form a salt bridge with the carboxylic group of the ligand, in a way that the substrate is orientating the alcohol group towards the catalytically relevant H467 and the FAD, helping the appropriate placement of the ligand for the oxidation. Both effects contribute to an activity improvement of the HMFO variant compared to the wild type.

The introduction of these mutations was tried in AAO. However, there are significant structural differences between both enzymes, being the size of the active site the main one. HMFO has a much larger cavity and the introduction of a bulky residue as an

arginine does not compromise the substrate diffusion. However, AAO is much smaller, and such residue does not even fit.

1.3. Laccases

Because of the expertise obtained by the group in laccases during the INDOX project (Santiago et al. 2016; Monza et al. 2015), the possibility to engineer a new laccase was offered. In this case though, a mediator was involved in the laccase catalysed oxidation. The initial idea was to study the bimolecular system, that is, the interaction between the mediator and the enzyme, in order to know the preferred binding spots. For this, docking techniques were used to sample the whole surface of LAC3. Interestingly, two main binding sites were spotted. One of them was close two the T1, while the other was on the opposite site of the protein, right next to the TNC (Figure 34).

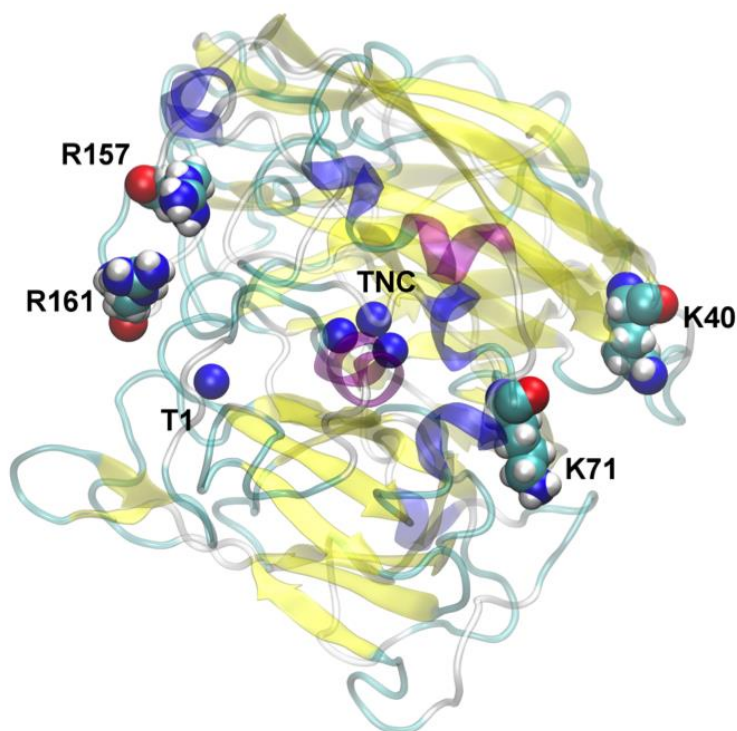


Figure 34. Laccase structure, where relevant ions and residues are highlighted (from Robert et al. 2017).

Electronic couplings have been estimated using the VMD Pathways plugin (Balabin et al. 2012). In this model, the coupling is quantified following continuous paths between the donor (a pyrene atom from the aromatic rings, in contact with the CNT) and the acceptor (each Cu atom). The model considers electron tunnelling through space, hydrogen bonds or covalent bonds, all characterized by different tunnelling barriers. In a simplified way, the electron coupling value is determined by the number and nature of tunnelling events.

Simple electron transfer calculations were measured between Ru and the T1 copper, which is the first electron acceptor, by quantifying continuous paths between both atoms using the VMD Pathways plugin (Balabin et al. 2012). This model determines the electron coupling value by the number and nature of tunnelling events, discerning between through space, hydrogen bonds or covalent bonds (each with a different tunnelling barrier). This method will be referred as “classic electron transfer” calculations. As expected, the results showed much better electron coupling when the ruthenium complex is in the T1 pocket, suggesting that electron transfer occurs mostly through this conformation.

Intuitively, covalently binding the photosensitizer to the protein should increase the oxidative activity of the whole system, since the mediator would be fixed in a position. Moreover, since experimental conditions do not work under saturation concentration of $[\text{Ru}(\text{bpy})_3]^{2+}$, grafting would also solve diffusion-related limitations. Therefore, different unimolecular systems (LAC3-Ru, UNIK157-Ru and UNIK-161) were modelled and, once again, classic electron transfer calculations were simulated in order to estimate the electron coupling between the donor (ruthenium) and the acceptor (T1 copper).

Results proved that, after grafting, the electron coupling decreased by more than two orders of magnitude for UNIK161 and UNIK157 (the ones close to the T1). For the LAC3-Ru, where the photosensitizer is grafted to either Lys40 or Lys71 (present in the native LAC3 and located in the TNC pocket) electron coupling values towards the T1 decrease are half of those for the UNIK variants. However, a direct electron transfer to the TNC was also considered for the different systems (Table 5).

Table 5. Calculated TDA values between Ru and Cu ions for LAC3-Ru, UNIK₁₆₁-Ru and UNIK₁₅₇-Ru.

System	$T_{\text{DA}}^{(\text{Ru},\text{T1})}$ (eV)	$T_{\text{DA}}^{(\text{Ru},\text{TNC})}$ (eV)
UNIK157-Ru	1×10^{-7}	1×10^{-7}
UNIK161-Ru	1×10^{-7}	5×10^{-8}
LAC3-Ru (K40)	2×10^{-13}	2×10^{-10}
LAC3-Ru (K71)	6×10^{-11}	3×10^{-9}

The most stable position of the ruthenium complex was revealed through metadynamics studies, indicating that $[\text{Ru}(\text{bpy})_3]^{2+}$ preferable position is outside the TNC cavity, for UNIK71 and UNIK40 systems (Figure 35). Therefore, electron transfer simulations were performed from these positions. Notice that, for the sake of simplicity, the contribution of both lysines is evaluated separately in the LAC3-Ru complex. Interestingly, this revealed that the electron coupling from Lys71 is better than Lys40, suggesting that UNIK71-Ru would already outperform the LAC3-Ru system.

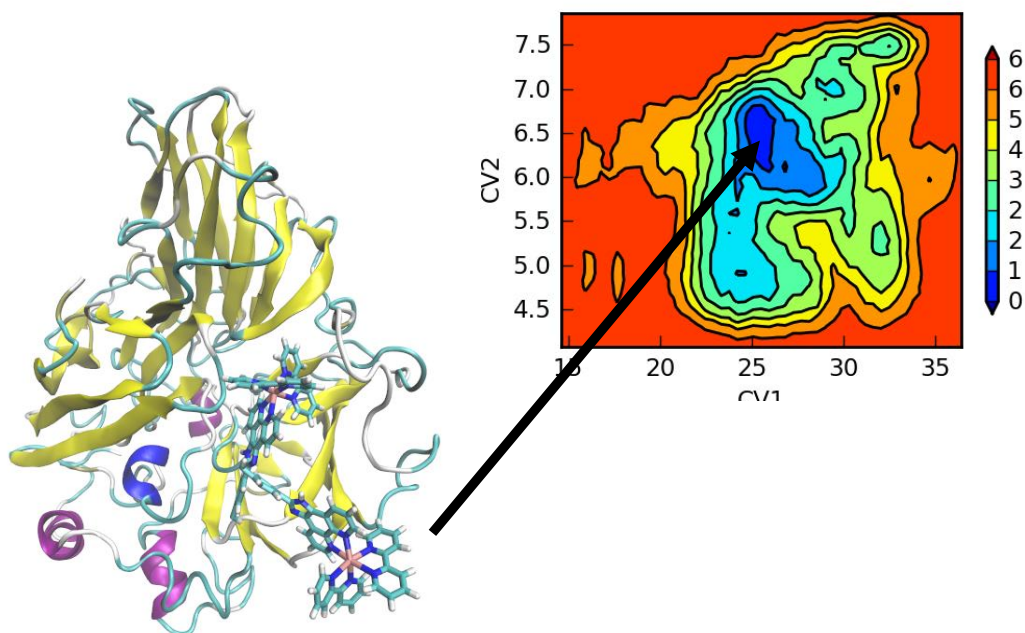


Figure 35. Result of the metadynamics simulation of UNIK40-Ru. CV1 is the Ru-T2 copper distance, CV2 is the distance between the N atom of K40 and its alpha carbon atom. Similar results are obtained for the UNIK71-Ru variant.

Results suggested that, according to the obtained electron coupling values, a direct electron transfer to the TNC was, not only possible, but even more favoured for the LAC3-Ru system. Therefore, since the T1 copper is no longer needed because its function is performed by the ruthenium, a possible way to improve the activity of the enzyme would be to maximize the electron transfer to the TNC, in order to avoid the internal electron transfer, and implicitly, the back-transfer problem. Therefore, the goal was to find the best position over the protein surface where the ruthenium complex could be attached. In order to do so, a LAC3 variant with no lysines (K0) was modelled. This included mutations K40S and K71S, where serine was chosen to preserve some polarity. From K0 variant, all residues present in the protein surface were mutated to lysine, one by one, and then the ruthenium complex was attached. After the relaxation of the system,

the different positions were scored according to the electron coupling values, and the results are compiled in Table 6.

Table 6. Calculated TDA values between Ru and TNC relative to UNIK71-Ru

System	$T_{DA}^{(Ru,TNC)}$ (eV)
UNIK83-Ru	100
UNIK349-Ru	50
UNIK157-Ru	33
UNIK75-Ru	25
UNIK69-Ru	20
UNIK161_C453S-Ru	17
UNIK76-Ru	15
UNIK73-Ru	10
UNIK405-Ru	10

Since UNIK71-Ru would perform better than LAC3-Ru, as it was concluded earlier, the electron coupling for this system was taken as a reference value. According to these results, by aiming for the best electron transfer directionality towards the TNC, electron coupling values could be improved from 10 to 100 times, meaning that electron transfer would improve from 100 to 10000 times, following the Marcus equation. Ideally, these variants should be tested experimentally, removing the T1 copper from the laccase. The position of these residues in the protein are illustrated in Figure 36. Notice that, for UNIK161, the mutation C453S is included, which removes the T1 copper. This is specified because, otherwise, the electron transfer to the T1 would be faster.

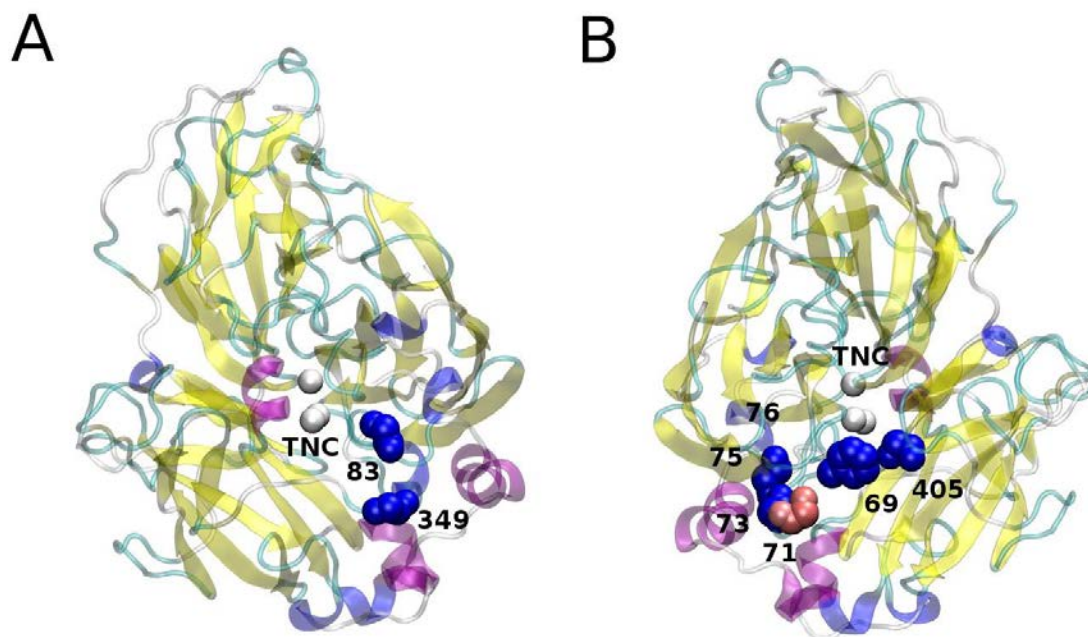


Figure 36 Result of the metadynamics simulation of UNIK40-Ru. CV1 is the Ru-T2 copper distance, CV2 is the distance between the N atom of K40 and its alpha carbon atom. Similar results are obtained for the UNIK71-Ru variant.

The main difference between these variants and UNIK157 or UNIK161 is that the latter are not designed with a purpose, but simply substituted from arginines, in positions close to the T1, in order to not disturb the structure of the protein. However, these new variants aim for the shortest path for donor-acceptor electron transfer. Even though this study is pending of experimental validation, preliminary results seem to indicate that UNIK71 variant perform even better than any previous variant tested so far.

1.3.1. Interaction of laccases with a nanotube

From this previous study, where UNIK71 was found to be a promising variant for direct electron transfer to the TNC, new experiments were performed in the lab. Then, the idea of simulating the interaction of UNIK161, UNIK40 and UNIK71 with a 100Å-diameter carbon nanotube (CNT) was suggested (Figure 37), to check if the same electron coupling results would be obtained, or if it would be affected by the structural modification of the protein due to the interaction with the nanotube. In this case, instead of grafting to a lysine a ruthenium complex, a ligand with a π -system group is added to the protein, so it can bind to the CNT by a strong π - π stacking interaction. After docking (to check the preferred orientation of the protein) and MD simulations (to relax the system), classic

electron transfer calculations were performed between the different copper ions and the closest carbon atom of the nanotube.

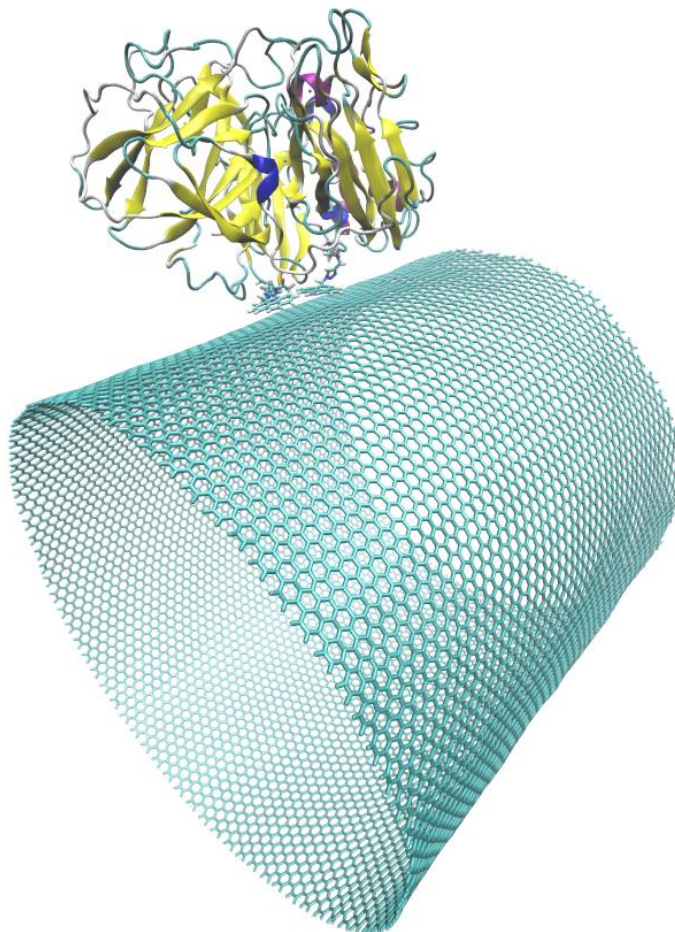


Figure 37. Representation of the modelled system, with LAC3 bound to the CNT through pi-stacking interactions.

In agreement with previous calculations, UNIK71 electron transfer is much more efficient than UNIK40. Between UNIK161 and UNIK71, no significant differences in electron coupling values were obtained, although T1 seems to be favoured by UNIK161 while UNIK71 transfers the electron directly to the T3. Results are presented in Table 7, and differences between electron transfer calculations with the CNT electrode and the Ru-complex are shown in Table 8.

Table 7. Calculated TDA values between the CNT electrode and Cu ions from different positions of the protein.

System	$T_{DA}(T1)$ (eV)	$T_{DA}(T2)$ (eV)	$T_{DA}(T3)$ (eV)	$T_{DA}(T3')$ (eV)
UNIK40	3×10^{-13}	5×10^{-11}	2×10^{-11}	1×10^{-10}
UNIK71	5×10^{-11}	2×10^{-9}	2×10^{-8}	3×10^{-9}
UNIK161	3×10^{-08}	2×10^{-11}	1×10^{-9}	7×10^{-10}

Table 8. Comparison of calculated TDA values to the CNT vs the Ru complex.

Cu	$T_{DA}^{(Cu, CNT)}$ (eV)	$T_{DA}^{(Cu, Ru)}$ (eV)
T1 from UNIK161	3×10^{-8}	5×10^{-7}
T2 from UNIK71	2×10^{-9}	1×10^{-8}
T3 from UNIK71	2×10^{-8}	4×10^{-9}
T3' from UNIK71	3×10^{-9}	1×10^{-8}

From these simulations, it seems clear that UNIK71 variant should be tested in the lab with the CNT too. In fact, this was already suggested to the experimental collaborators, and it was seen that, in fact, UNIK71 perform even better than UNIK161. These results are not included as a paper in this thesis, although the manuscript is pending of being submitted.

V. CONCLUSIONS

The conclusion that can be extracted from this thesis are:

1. Computational studies are shown to be a useful complement to experimental work for enzyme engineering. These methods can provide, not only insights of the reaction mechanism or the role of certain amino acids, but they also are valid approaches to design and predict improved protein variants.
2. When collaborating with experimental partners and industries, focusing on detailed predictions may not be as useful as more general predictions, as reflected in our flavoprotein designs. Despite the efforts to suggest precise mutations, a contact map identified positions 500 and 501 for aryl-alcohol oxidase, for which combinatorial mutagenesis resulted in improved variants for secondary benzyl-alcohols.
3. Protein-ligand interactions have been characterised for all the steps of the 5-hydroxymethylfurfural oxidation chain in aryl-alcohol oxidase and 5-hydroxymethylfurfural oxidase. Moreover, a negative charge was identified as the main reason for non-reactivity with aryl-alcohol oxidase. However, an enhanced variant for this enzyme could not be obtained, although we took one step towards it.
4. The use of a photosensitizer as a mediator with laccases open the possibility to transfer the electrons directly to the tri-nuclear cluster. Therefore, the engineering of laccases should be focused on maximizing the electron transfer directionality there, to increase the reactivity.

VI. REFERENCES

- Allen, Benjamin D., Alex Nisthal, and Stephen L. Mayo. 2010. "Experimental Library Screening Demonstrates the Successful Application of Computational Protein Design to Large Structural Ensembles." *Proceedings of the National Academy of Sciences of the United States of America* 107 (46): 19838–43.
- AlQuraishi, Mohammed. 2019. "AlphaFold at CASP13." *Bioinformatics*, May. <https://doi.org/10.1093/bioinformatics/btz422>.
- Amidon, Thomas E., Christopher D. Wood, Alan M. Shupe, Yang Wang, Mitchell Graves, and Shijie Liu. 2008. "Biorefinery: Conversion of Woody Biomass to Chemicals, Energy and Materials." *Journal of Biobased Materials and Bioenergy* 2 (2): 100–120.
- Anwar, Jamshed, and Dirk Zahn. 2017. "Polymorphic Phase Transitions: Macroscopic Theory and Molecular Simulation." *Advanced Drug Delivery Reviews* 117 (August): 47–70.
- Arnold, Frances H., Keqin Chen, Chara Economou, Wayne Chen, Pascal Martinez, Kyung Pyo Yoon, and Mariana Van Dam. 1993. "Engineering Nonaqueous Solvent-Compatible Enzymes." *ACS Symposium Series*. <https://doi.org/10.1021/bk-1993-0516.ch008>.
- Asgher, Muhammad, Muhammad Shahid, Shagufta Kamal, and Hafiz Muhammad Nasir Iqbal. 2014. "Recent Trends and Valorization of Immobilization Strategies and Lignolytic Enzymes by Industrial Biotechnology." *Journal of Molecular Catalysis B: Enzymatic*. <https://doi.org/10.1016/j.molcatb.2013.12.016>.
- Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. 2001. "Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model." *Biophysical Journal* 80 (1): 505–15.
- Atkins, Peter W., and Ronald S. Friedman. 2011. *Molecular Quantum Mechanics*. OUP Oxford.
- Bader, Richard F. W. 1994. *Atoms in Molecules: A Quantum Theory*. Clarendon Press.
- Bajaj, M. 1984. "Evolution and the Tertiary Structure of Proteins." *Annual Review of Biophysics and Biomolecular Structure*. <https://doi.org/10.1146/annurev.biophys.13.1.453>.
- Baker, Christopher M. 2015. "Polarizable Force Fields for Molecular Dynamics Simulations of Biomolecules." *Wiley Interdisciplinary Reviews: Computational Molecular Science* 5 (2): 241–54.
- Baker Dockrey, Summer A., and Alison R. H. Narayan. 2019. "Flavin-Dependent Biocatalysts in Synthesis." *Tetrahedron* 75 (9): 1115–21.
- Balabin, Ilya A., Xiangqian Hu, and David N. Beratan. 2012. "Exploring Biological Electron Transfer Pathway Dynamics with the Pathways Plugin for VMD." *Journal of Computational Chemistry*. <https://doi.org/10.1002/jcc.22927>.

- Bar-Even, Arren, Ron Milo, Elad Noor, and Dan S. Tawfik. 2015. "The Moderately Efficient Enzyme: Futile Encounters and Enzyme Floppiness." *Biochemistry* 54 (32): 4969–77.
- Bar-Even, Arren, Elad Noor, Yonatan Savir, Wolfram Liebermeister, Dan Davidi, Dan S. Tawfik, and Ron Milo. 2011. "The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters." *Biochemistry* 50 (21): 4402–10.
- Barrozo, Alexandre, Rok Borstnar, Gaël Marloie, and Shina Caroline Lynn Kamerlin. 2012. "Computational Protein Engineering: Bridging the Gap between Rational Design and Laboratory Evolution." *International Journal of Molecular Sciences* 13 (10): 12428–60.
- Berraud-Pache, Romain, Cristina Garcia-Iriepa, and Isabelle Navizet. 2018. "Modeling Chemical Reactions by QM/MM Calculations: The Case of the Tautomerization in Fireflies Bioluminescent Systems." *Frontiers in Chemistry*. <https://doi.org/10.3389/fchem.2018.00116>.
- Borrelli, Kenneth W., Benjamin Cossins, and Victor Guallar. 2010. "Exploring Hierarchical Refinement Techniques for Induced Fit Docking with Protein and Ligand Flexibility." *Journal of Computational Chemistry* 31 (6): 1224–35.
- Borrelli, Kenneth W., Andreas Vitalis, Raul Alcantara, and Victor Guallar. 2005. "PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique." *Journal of Chemical Theory and Computation*. <https://doi.org/10.1021/ct0501811>.
- Bujanovic, Biljana M., Mangesh J. Goundalkar, and Thomas E. Amidon. 2012. "Increasing the Value of a Biorefinery Based on Hot-Water Extraction: Lignin Products." *Tappi Journal* 11 (1): 19–26.
- Cadwell, R. C., and G. F. Joyce. 1992. "Randomization of Genes by PCR Mutagenesis." *PCR Methods and Applications* 2 (1): 28–33.
- Caldararu, Octav, Milica Feldt, Daniela Cioloboc, Marie-Céline van Severen, Kerstin Starke, Ricardo A. Mata, Ebbe Nordlander, and Ulf Ryde. 2018. "QM/MM Study of the Reaction Mechanism of Sulfite Oxidase." *Scientific Reports* 8 (1): 4684.
- Chapman, Jordan, Ahmed Ismail, and Cerasela Dinu. 2018. "Industrial Applications of Enzymes: Recent Advances, Techniques, and Outlooks." *Catalysts*. <https://doi.org/10.3390/catal8060238>.
- Chica, Roberto A., Nicolas Doucet, and Joelle N. Pelletier. 2005. "Semi-Rational Approaches to Engineering Enzyme Activity: Combining the Benefits of Directed Evolution and Rational Design." *Current Opinion in Biotechnology* 16 (4): 378–84.
- Cohen, S. N., A. C. Chang, H. W. Boyer, and R. B. Helling. 1973. "Construction of Biologically Functional Bacterial Plasmids in Vitro." *Proceedings of the National Academy of Sciences of the United States of America* 70 (11): 3240–44.

- Cossins, Benjamin P., Ali Hosseini, and Victor Guallar. 2012. "Exploration of Protein Conformational Change with PELE and Meta-Dynamics." *Journal of Chemical Theory and Computation* 8 (3): 959–65.
- Cramer, Christopher J., and Donald G. Truhlar. 1999. "Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics." *Chemical Reviews*. <https://doi.org/10.1021/cr960149m>.
- Darby, John F., Masakazu Atobe, James D. Firth, Paul Bond, Gideon J. Davies, Peter O'Brien, and Roderick E. Hubbard. 2017. "Increase of Enzyme Activity through Specific Covalent Modification with Fragments." *Chemical Science* 8 (11): 7772–79.
- DeGrado, W. F., F. G. Prendergast, H. R. Wolfe Jr, and J. A. Cox. 1985. "The Design, Synthesis, and Characterization of Tight-Binding Inhibitors of Calmodulin." *Journal of Cellular Biochemistry* 29 (2): 83–93.
- Dijkman, Willem P., Claudia Binda, Marco W. Fraaije, and Andrea Mattevi. 2015. "Structure-Based Enzyme Tailoring of 5-Hydroxymethylfurfural Oxidase." *ACS Catalysis* 5 (3): 1833–39.
- Dourado, Daniel F. A. R., Marcel Swart, and Alexandra T. P. Carvalho. 2018. "Why the Flavin Adenine Dinucleotide (FAD) Cofactor Needs To Be Covalently Linked to Complex II of the Electron-Transport Chain for the Conversion of FADH into FAD." *Chemistry* 24 (20): 5246–52.
- Durão, Paulo, Isabel Bento, André T. Fernandes, Eduardo P. Melo, Peter F. Lindley, and Lígia O. Martins. 2006. "Perturbations of the T1 Copper Site in the CotA Laccase from *Bacillus Subtilis*: Structural, Biochemical, Enzymatic and Stability Studies." *Journal of Biological Inorganic Chemistry: JBIC: A Publication of the Society of Biological Inorganic Chemistry* 11 (4): 514.
- Edwards, Ana Maria. 2014. "Structure and General Properties of Flavins." *Methods in Molecular Biology* 1146: 3–13.
- Estell, D. A., T. P. Graycar, and J. A. Wells. 1985. "Engineering an Enzyme by Site-Directed Mutagenesis to Be Resistant to Chemical Oxidation." *The Journal of Biological Chemistry* 260 (11): 6518–21.
- Fei, Qiang, Michael T. Guarnieri, Ling Tao, Lieve M. L. Laurens, Nancy Dowe, and Philip T. Pienkos. 2014. "Bioconversion of Natural Gas to Liquid Fuel: Opportunities and Challenges." *Biotechnology Advances* 32 (3): 596–614.
- Ferreira, P., A. Hernandez-Ortega, and B. Herguedas. 2009. "Aryl-Alcohol Oxidase Involved in Lignin Degradation A MECHANISTIC STUDY BASED ON STEADY AND PRE-STEADY STATE KINETICS AND PRIMARY AND" *Journal of Biological*. <http://www.jbc.org/content/284/37/24840.short>.
- Fox, Richard J., and Gjalte W. Huisman. 2008. "Enzyme Optimization: Moving from Blind Evolution to Statistical Exploration of Sequence–function Space." *Trends in Biotechnology*. <https://doi.org/10.1016/j.tibtech.2007.12.001>.

- Fraaije, M. W., R. H. van den Heuvel, W. J. van Berkel, and A. Mattevi. 1999. "Covalent Flavinylation Is Essential for Efficient Redox Catalysis in Vanillyl-Alcohol Oxidase." *The Journal of Biological Chemistry* 274 (50): 35514–20.
- Ghisla, S., and V. Massey. 1989. "Mechanisms of Flavoprotein-Catalyzed Reactions." *European Journal of Biochemistry / FEBS* 181 (1): 1–17.
- Goeddel, D. V., D. G. Kleid, F. Bolivar, H. L. Heyneker, D. G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura, and A. D. Riggs. 1979. "Expression in *Escherichia Coli* of Chemically Synthesized Genes for Human Insulin." *Proceedings of the National Academy of Sciences of the United States of America* 76 (1): 106–10.
- Goldsmith, Moshe, and Dan S. Tawfik. 2017. "Enzyme Engineering: Reaching the Maximal Catalytic Efficiency Peak." *Current Opinion in Structural Biology* 47 (December): 140–50.
- Gustafsson, Claes, Sridhar Govindarajan, and Jeremy Minshull. 2003. "Putting Engineering Back into Protein Engineering: Bioinformatic Approaches to Catalyst Design." *Current Opinion in Biotechnology* 14 (4): 366–70.
- Guvench, Olgun, and Alexander D. MacKerell Jr. 2008. "Comparison of Protein Force Fields for Molecular Dynamics Simulations." *Methods in Molecular Biology* 443: 63–88.
- Hagen, Joel B. 2000. "The Origins of Bioinformatics." *Nature Reviews Genetics*. <https://doi.org/10.1038/35042090>.
- Hart, Kathryn M., Chris M. W. Ho, Supratik Dutta, Michael L. Gross, and Gregory R. Bowman. 2016. "Modelling Proteins' Hidden Conformations to Predict Antibiotic Resistance." *Nature Communications*. <https://doi.org/10.1038/ncomms12965>.
- Hernández-Ortega, Aitor, Patricia Ferreira, and Angel T. Martínez. 2012. "Fungal Aryl-Alcohol Oxidase: A Peroxide-Producing Flavoenzyme Involved in Lignin Degradation." *Applied Microbiology and Biotechnology* 93 (4): 1395–1410.
- Hernández-Ortega, Aitor, Patricia Ferreira, Pedro Merino, Milagros Medina, Victor Guallar, and Angel T. Martínez. 2012. "Stereoselective Hydride Transfer by Aryl-Alcohol Oxidase, a Member of the GMC Superfamily." *Chembiochem: A European Journal of Chemical Biology* 13 (3): 427–35.
- Hernández-Ortega, Aitor, Fátima Lucas, Patricia Ferreira, Milagros Medina, Victor Guallar, and Angel T. Martínez. 2011. "Modulating O₂ Reactivity in a Fungal Flavoenzyme." *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.m111.282467>.
- Hosseini, Ali, Moran Brouk, Maria Fatima Lucas, Fabian Glaser, Ayelet Fishman, and Victor Guallar. 2015. "Atomic Picture of Ligand Migration in Toluene 4-Monooxygenase." *The Journal of Physical Chemistry. B* 119 (3): 671–78.
- Huang, Sheng-You, and Xiaoqin Zou. 2007. "Ensemble Docking of Multiple Protein Structures: Considering Protein Structural Variations in Molecular Docking." *Proteins* 66 (2): 399–421.

- Huisman, Gjalb W., and Steven J. Collier. 2013. "On the Development of New Biocatalytic Processes for Practical Pharmaceutical Synthesis." *Current Opinion in Chemical Biology* 17 (2): 284–92.
- Jacobson, Matthew P., George A. Kaminski, Richard A. Friesner, and Chaya S. Rapp. 2002. "Force Field Validation Using Protein Side Chain Prediction." *The Journal of Physical Chemistry B*. <https://doi.org/10.1021/jp021564n>.
- Jemli, Sonia, Dorra Ayadi-Zouari, Hajer Ben Hlima, and Samir Bejar. 2016. "Biocatalysts: Application and Engineering for Industrial Purposes." *Critical Reviews in Biotechnology* 36 (2): 246–58.
- Jiang, Lin, Eric A. Althoff, Fernando R. Clemente, Lindsey Doyle, Daniela Röthlisberger, Alexandre Zanghellini, Jasmine L. Gallaher, et al. 2008. "De Novo Computational Design of Retro-Aldol Enzymes." *Science* 319 (5868): 1387–91.
- Jortzik, Esther, Lihui Wang, Jipeng Ma, and Katja Becker. 2014. "Flavins and Flavoproteins: Applications in Medicine." *Methods in Molecular Biology* 1146: 113–57.
- Juris, A., and L. Moggi. 1981. "Kinetics of the Catalysed Oxidation of H₂O by Ru(bipy)₃²⁺." *Journal of Photochemistry*. [https://doi.org/10.1016/0047-2670\(81\)85294-x](https://doi.org/10.1016/0047-2670(81)85294-x).
- Kamerlin, Shina C. L., and Arieh Warshel. 2011. "The Empirical Valence Bond Model: Theory and Applications." *Wiley Interdisciplinary Reviews. Computational Molecular Science* 1 (1): 30–45.
- Kapoor, Swati, Aasima Rafiq, and Savita Sharma. 2017. "Protein Engineering and Its Applications in Food Industry." *Critical Reviews in Food Science and Nutrition*. <https://doi.org/10.1080/10408398.2014.1000481>.
- Klonowska, Agnieszka, Christian Gaudin, Marcel Asso, André Fournel, Marius Réglie, and Thierry Tron. 2005. "LAC3, a New Low Redox Potential Laccase from *Trametes* Sp. Strain C30 Obtained as a Recombinant Protein in Yeast." *Enzyme and Microbial Technology*. <https://doi.org/10.1016/j.enzmictec.2004.03.022>.
- Koch, Wolfram, and Max C. Holthausen. 2015. *A Chemist's Guide to Density Functional Theory*. John Wiley & Sons.
- Kohn, W., and L. J. Sham. 1965. "Self-Consistent Equations Including Exchange and Correlation Effects." *Physics Review* 140 (4A): A1133–38.
- Kontoyianni, Maria. 2017. "Docking and Virtual Screening in Drug Discovery." *Methods in Molecular Biology* 1647: 255–66.
- Koopman, F., N. Wierckx, J. H. de Winde, and H. J. Ruijsenaars. 2010. "Identification and Characterization of the Furfural and 5-(hydroxymethyl)furfural Degradation Pathways of *Cupriavidus Basilensis* HMF14." *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.0913039107>.

- Korendovych, Ivan V. 2018. "Rational and Semirational Protein Design." *Methods in Molecular Biology* 1685: 15–23.
- Kunamneni, Adinarayana, Francisco J. Plou, Antonio Ballesteros, and Miguel Alcalde. 2008. "Laccases and Their Applications: A Patent Review." *Recent Patents on Biotechnology* 2 (1): 10–24.
- Laio, Alessandro, and Michele Parrinello. 2002. "Escaping Free-Energy Minima." *Proceedings of the National Academy of Sciences of the United States of America* 99 (20): 12562–66.
- Leach, A. 2001. "Empirical Force Field Models: Molecular Mechanics." A. Leach, *Molecular Modelling: Principles and Applications*, 165–252.
- Lecina, Daniel, Joan F. Gilabert, and Victor Guallar. 2017. "Adaptive Simulations, towards Interactive Protein-Ligand Modeling." *Scientific Reports* 7 (1): 8466.
- Lewis, Charles A., Jr, and Richard Wolfenden. 2008. "Uroporphyrinogen Decarboxylation as a Benchmark for the Catalytic Proficiency of Enzymes." *Proceedings of the National Academy of Sciences of the United States of America* 105 (45): 17328–33.
- Li, Shuang, Xiaofeng Yang, Shuai Yang, Muzi Zhu, and Xiaoning Wang. 2012. "Technology Prospecting on Enzymes: Application, Marketing and Engineering." *Computational and Structural Biotechnology Journal* 2 (November): e201209017.
- Mack, Korrie L., and James Shorter. 2016. "Engineering and Evolution of Molecular Chaperones and Protein Disaggregases with Enhanced Activity." *Frontiers in Molecular Biosciences* 3 (March): 8.
- Madadkar-Sobhani, Armin, and Victor Guallar. 2013. "PELE Web Server: Atomistic Study of Biomolecular Systems at Your Fingertips." *Nucleic Acids Research* 41 (Web Server issue): W322–28.
- Marcus, R. A. 1956. "On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer. I." *The Journal of Chemical Physics* 24 (5): 966–78.
- Marques, Daniela Viana, Suellen Machado, Valéria Ebinuma, Carolina Duarte, Attilio Converti, and Ana Porto. 2018. "Production of β -Lactamase Inhibitors by *Streptomyces* Species." *Antibiotics*. <https://doi.org/10.3390/antibiotics7030061>.
- Martin, Caterina, Amaury Ovalle Maqueo, Hein J. Wijma, and Marco W. Fraaije. 2018. "Creating a More Robust 5-Hydroxymethylfurfural Oxidase by Combining Computational Predictions with a Novel Effective Library Design." *Biotechnology for Biofuels* 11 (March): 56.
- Mateo, Cesar, Jose M. Palomo, Gloria Fernandez-Lorente, Jose M. Guisan, and Roberto Fernandez-Lafuente. 2007. "Improvement of Enzyme Activity, Stability and Selectivity via Immobilization Techniques." *Enzyme and Microbial Technology*. <https://doi.org/10.1016/j.enzmictec.2007.01.018>.

- Mazzotta, Cosimo, Stefano Caragiuli, and Aldo Caporossi. 2014. "Riboflavin and the Cornea and Implications for Cataracts." In *Handbook of Nutrition, Diet and the Eye*, 123–30. Elsevier.
- Michaelis, Leonor, Maud Leonora Menten, Kenneth A. Johnson, and Roger S. Goody. 2011. "The Original Michaelis Constant: Translation of the 1913 Michaelis-Menten Paper." *Biochemistry* 50 (39): 8264–69.
- Miller, Brian G., and Richard Wolfenden. 2002. "Catalytic Proficiency: The Unusual Case of OMP Decarboxylase." *Annual Review of Biochemistry* 71: 847–85.
- Mohamad, Nur Royhaila, Nur Haziqah Che Marzuki, Nor Aziah Buang, Fahrul Huyop, and Roswanira Abdul Wahab. 2015. "An Overview of Technologies for Immobilization of Enzymes and Surface Analysis Techniques for Immobilized Enzymes." *Biotechnology, Biotechnological Equipment* 29 (2): 205–20.
- Monza, Emanuele, M. Fatima Lucas, Susana Camarero, Lorea C. Alejaldre, Angel T. Martínez, and Victor Guallar. 2015. "Insights into Laccase Engineering from Molecular Simulations: Toward a Binding-Focused Strategy." *The Journal of Physical Chemistry Letters*. <https://doi.org/10.1021/acs.jpcllett.5b00225>.
- Morozova, O. V., G. P. Shumakovich, M. A. Gorbacheva, S. V. Shleev, and A. I. Yaropolov. 2007. "'Blue' Laccases." *Biochemistry* 72 (10): 1136–50.
- Moult, John, Krzysztof Fidelis, Andriy Kryshtafovych, Torsten Schwede, and Anna Tramontano. 2018. "Critical Assessment of Methods of Protein Structure Prediction (CASP)-Round XII." *Proteins* 86 Suppl 1 (March): 7–15.
- Mulliken, R. S. 1955. "Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I." *The Journal of Chemical Physics* 23 (10): 1833–40.
- Nelson, David L., Albert L. Lehninger, and Michael M. Cox. 2008. *Lehninger Principles of Biochemistry*. Macmillan.
- O'Neil, K. T., and W. F. DeGrado. 1985. "A Predicted Structure of Calmodulin Suggests an Electrostatic Basis for Its Function." *Proceedings of the National Academy of Sciences of the United States of America* 82 (15): 4954–58.
- Ott, J. Bevan, J. Bevan Ott, and Juliana Boerio-Goates. 2000. "Summary of Thermodynamic Relationships." *Chemical Thermodynamics: Advanced Applications*. <https://doi.org/10.1016/b978-012530985-1.50002-3>.
- Pickl, Mathias, Alexander Swoboda, Elvira Romero, Christoph K. Winkler, Claudia Binda, Andrea Mattevi, Kurt Faber, and Marco W. Fraaije. 2018. "Kinetic Resolution of Sec-Thiols by Enantioselective Oxidation with Rationally Engineered 5-(Hydroxymethyl) Furfural Oxidase." *Angewandte Chemie* 130 (11): 2914–18.
- Radzicka, A., and R. Wolfenden. 1995. "A Proficient Enzyme." *Science*. <https://doi.org/10.1126/science.7809611>.

- Reznik, Ed, Pankaj Mehta, and Daniel Segrè. 2013. "Flux Imbalance Analysis and the Sensitivity of Cellular Growth to Changes in Metabolite Pools." *PLoS Computational Biology* 9 (8): e1003195.
- Riva, Sergio. 2006. "Laccases: Blue Enzymes for Green Chemistry." *Trends in Biotechnology* 24 (5): 219–26.
- Robert, Viviane, Emanuele Monza, Lionel Tarrago, Ferran Sancho, Anna De Falco, Ludovic Schneider, Eloïne Npetgat Ngoutane, et al. 2017. "Probing the Surface of a Laccase for Clues towards the Design of Chemo-Enzymatic Catalysts." *ChemPlusChem*. <https://doi.org/10.1002/cplu.201700030>.
- Romero, Elvira, J. Rubén Gómez Castellanos, Giovanni Gadda, Marco W. Fraaije, and Andrea Mattevi. 2018. "Same Substrate, Many Reactions: Oxygen Activation in Flavoenzymes." *Chemical Reviews* 118 (4): 1742–69.
- Romero-Rivera, Adrian, Marc Garcia-Borràs, and Sílvia Osuna. 2016. "Computational Tools for the Evaluation of Laboratory-Engineered Biocatalysts." *Chemical Communications* 53 (2): 284–97.
- Röthlisberger, Daniela, Olga Khersonsky, Andrew M. Wollacott, Lin Jiang, Jason DeChancie, Jamie Betker, Jasmine L. Gallaher, et al. 2008. "Kemp Elimination Catalysts by Computational Enzyme Design." *Nature* 453 (7192): 190–95.
- Saio, Tomohide, Soichiro Kawagoe, Koichiro Ishimori, and Charalampos G. Kalodimos. 2018. "Oligomerization of a Molecular Chaperone Modulates Its Activity." *eLife* 7 (May). <https://doi.org/10.7554/eLife.35731>.
- Sajitz-Hermstein, Max, and Zoran Nikoloski. 2016. "Multi-Objective Shadow Prices Point at Principles of Metabolic Regulation." *Bio Systems* 146 (August): 91–101.
- Santiago, Gerard, Felipe de Salas, M. Fátima Lucas, Emanuele Monza, Sandra Acebes, Ángel T. Martínez, Susana Camarero, and Víctor Guallar. 2016. "Computer-Aided Laccase Engineering: Toward Biological Oxidation of Arylamines." *ACS Catalysis*. <https://doi.org/10.1021/acscatal.6b01460>.
- Schlick, Tamar. 2010. *Molecular Modeling and Simulation: An Interdisciplinary Guide: An Interdisciplinary Guide*. Springer Science & Business Media.
- Schneider, Ludovic, Yasmina Mekmouche, Pierre Rousselot-Pailley, A. Jalila Simaan, Viviane Robert, Marius Réglie, Ally Aukauloo, and Thierry Tron. 2015. "Visible-Light-Driven Oxidation of Organic Substrates with Dioxygen Mediated by a [Ru(bpy)₃]²⁺/Laccase System." *ChemSusChem*. <https://doi.org/10.1002/cssc.201500602>.
- Serrano, Ana, Ferran Sancho, Javier Viña-González, Juan Carro, Miguel Alcalde, Victor Guallar, and Angel T. Martínez. 2019. "Switching the Substrate Preference of Fungal Aryl-Alcohol Oxidase: Towards Stereoselective Oxidation of Secondary Benzyl Alcohols." *Catalysis Science & Technology* 9 (3): 833–41.

- Shaw, David E. 2009. "Anton: A Specialized Machine for Millisecond-Scale Molecular Dynamics Simulations of Proteins." 2009 19th IEEE Symposium on Computer Arithmetic. <https://doi.org/10.1109/arith.2009.33>.
- Siegel, Justin B., Alexandre Zanghellini, Helena M. Lovick, Gert Kiss, Abigail R. Lambert, Jennifer L. St Clair, Jasmine L. Gallaher, et al. 2010. "Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction." *Science* 329 (5989): 309–13.
- Simaan, A. Jalila, Yasmina Mekmouche, Christian Herrero, Pierre Moreno, Ally Aukauloo, Jacques A. Delaire, Marius Réglier, and Thierry Tron. 2011. "Photoinduced Multielectron Transfer to a Multicopper Oxidase Resulting in Dioxygen Reduction into Water." *Chemistry* 17 (42): 11743–46.
- Sonnenberg, A. S. M., M. H. M. Visser, B. Lavrijssen, J. W. Cone, and P. M. Hendrickx. 2016. "Evaluation of King Oyster Mushroom Strains (*Pleurotus Eryngii*) on Selective Lignin Degradation in Wheat Straw: An Update." 2016-4. Wageningen: Wageningen UR. <https://library.wur.nl/WebQuery/wurpubs/498330>.
- Steiner, Kerstin, and Helmut Schwab. 2012. "Recent Advances in Rational Approaches for Enzyme Engineering." *Computational and Structural Biotechnology Journal* 2 (October): e201209010.
- Sumanth, J. V., David R. Swanson, and And Hong Jiang. 2005. "Scheduling Many-Body Short Range MD Simulations on a Cluster of Workstations and Custom VLSI Hardware." In *High Performance Computing - HiPC 2004*, 166–75. Springer Berlin Heidelberg.
- Szabo, Attila, and Neil S. Ostlund. 2012. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Courier Corporation.
- Tadesse, Mahelet Aweke, Alessandro D'Annibale, Carlo Galli, Patrizia Gentili, and Federica Sergi. 2008. "An Assessment of the Relative Contributions of Redox and Steric Issues to Laccase Specificity towards Putative Substrates." *Organic & Biomolecular Chemistry* 6 (5): 868–78.
- Takahashi, Ryoji, Víctor A. Gil, and Victor Guallar. 2014. "Monte Carlo Free Ligand Diffusion with Markov State Model Analysis and Absolute Binding Free Energy Calculations." *Journal of Chemical Theory and Computation* 10 (1): 282–88.
- Tong, Joo Chuan, and Shoba Ranganathan. 2013. "Computational T Cell Vaccine Design." *Computer-Aided Vaccine Design*. <https://doi.org/10.1533/9781908818416.59>.
- Voigt, C. A., S. L. Mayo, F. H. Arnold, and Z. G. Wang. 2001. "Computationally Focusing the Directed Evolution of Proteins." *Journal of Cellular Biochemistry. Supplement* Suppl 37: 58–63.
- Warshel, A., and M. Levitt. 1976. "Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme." *Journal of Molecular Biology* 103 (2): 227–49.

- Wilkinson, A. J., A. R. Fersht, D. M. Blow, and G. Winter. 1983. "Site-Directed Mutagenesis as a Probe of Enzyme Structure and Catalysis: Tyrosyl-tRNA Synthetase Cysteine-35 to Glycine-35 Mutation." *Biochemistry* 22 (15): 3581–86.
- Winter, G., A. R. Fersht, A. J. Wilkinson, M. Zoller, and M. Smith. 1982. "Redesigning Enzyme Structure by Site-Directed Mutagenesis: Tyrosyl tRNA Synthetase and ATP Binding." *Nature* 299 (5885): 756–58.
- Xu, F. 1996. "Oxidation of Phenols, Anilines, and Benzenethiols by Fungal Laccases: Correlation between Activity and Redox Potentials as Well as Halide Inhibition." *Biochemistry* 35 (23): 7608–14.
- Xu, Mengang, and Markus A. Lill. 2013. "Induced Fit Docking, and the Use of QM/MM Methods in Docking." *Drug Discovery Today: Technologies*. <https://doi.org/10.1016/j.ddtec.2013.02.003>.
- Yilmazer, Nusret Duygu, and Martin Korth. 2015. "Enhanced Semiempirical QM Methods for Biomolecular Interactions." *Computational and Structural Biotechnology Journal* 13 (February): 169–75.

Cover image designed by starline & freepik

